Tutoriat PS 2 - Random Variables

Mihai Duzi, Lucan Cristian, Lumînăraru Ionut

Contents

1 Introduction

1.1 Types of random variables

There are two types of random variables: **Discrete** and **Continuous**. Both are used to assign a value or weight to events — from winning or losing a bet to the occurrence of a natural event like a tornado.

1.2 Terminology and Notation

Let's start with an exercise.

I roll two dice at the same time. What is the probability that their sum is 8?

$$\Omega = \{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}$$
$$A = \{(i, j) \mid (i, j) \in \Omega, i + j = 8\}$$
$$P(A) =?$$

Let's define a function

$$X(i, j) = i + j, \quad X : \Omega \to S, \text{ where } S = \{2, 3, \ldots, 12\}$$

Now,

$$A = \{(i, j) \mid X(i, j) = 8\}$$

We can write

$$P(A) = P(X = 8) = \frac{5}{36}$$

This function X is a **discrete random variable**. If S is at most countable, X is discrete; otherwise, X is continuous.

2 Discrete Random Variables

Beyond the already existing notation, we define the **probability mass function (pmf)**:

$$p : S \to [0, 1]$$

$$p(x) = P(X = x)$$

For our example above, $p(8) = P(X = 8) = \frac{5}{36}$. The pmf uniquely identifies the distribution of the random variable.

For our example:

$$X \sim \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}$$

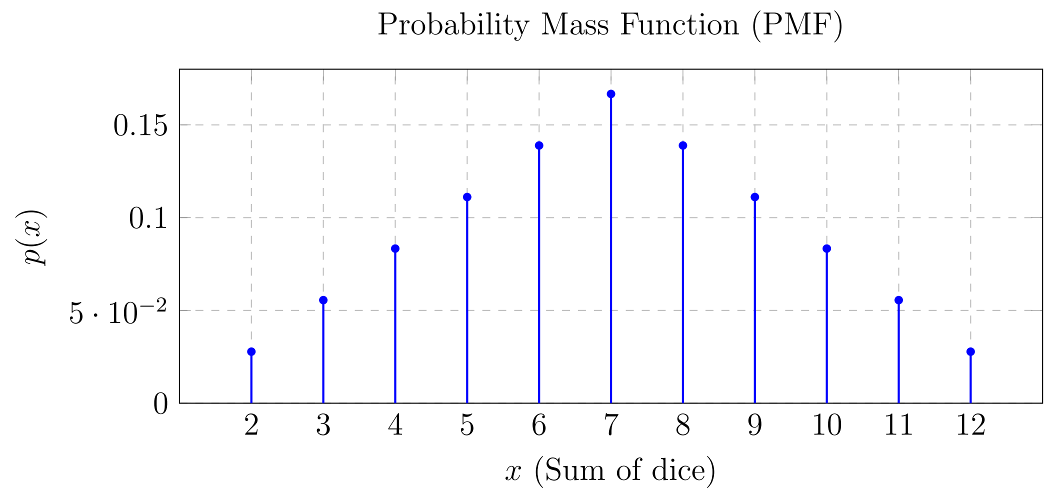
Now, what if I ask: what is the probability that the sum of the dice is smaller than 8?

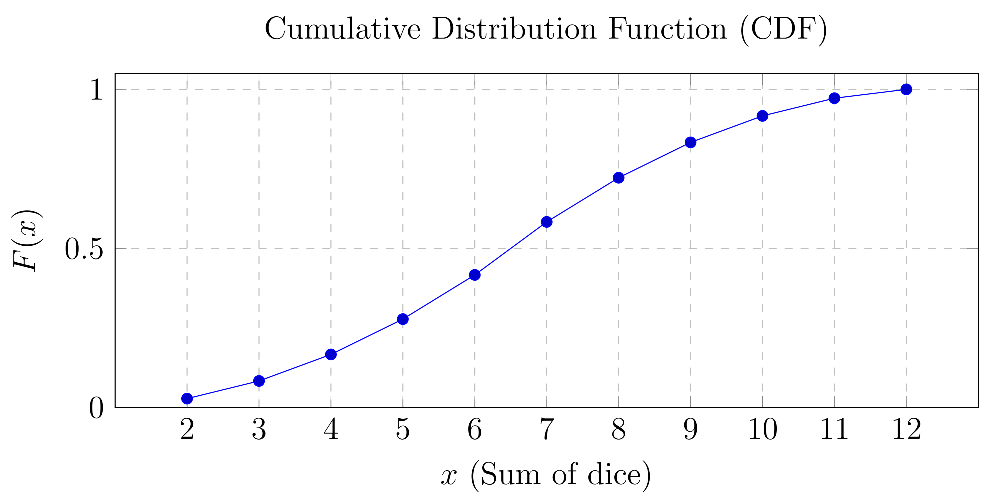
$$F(8) = P(X < 8) = \sum_{i=2}^{7} P(X = i)$$

This defines the **cumulative distribution function (CDF)**, which gives the probability that X takes a value less than or equal to a given number.

$$F(x) = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{3}{36} & \frac{6}{36} & \frac{10}{36} & \frac{15}{36} & \frac{21}{36} & \frac{26}{36} & \frac{30}{36} & \frac{33}{36} & \frac{35}{36} & 1 \end{pmatrix}$$

Visualizing the PMF and CDF





Observations

We can now observe the following properties of the CDF:

1. F is **non-decreasing**. That is, the graph never goes down — symbolically, if $a \leq b$, then $F(a) \leq F(b)$. 2. $0 \leq F(a) \leq 1$ for all real numbers a. 3. $\lim_{a \to -\infty} F(a) = 0$ and $\lim_{a \to \infty} F(a) = 1$.

Random Variables Arithmetic

"+", "-", "×" also work on random variables.

The following random variable represents the distribution of points of a football match and their probabilities (3- Victory, 1 - Draw, 0 - Loss).

$$X \sim \begin{pmatrix} 0 & 1 & 3 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

Now what if we want the distribution of points of playing two football matches? $Y = X_1 + X_2 = 2 \times X$

Probability table:

$X_1 \backslash X_2$	0	1	3
0	0	1	3
1	1	2	4
3	3	4	6

Each combination occurs with probability equal to the product of the two individual probabilities:

$$P(X_1 = i, X_2 = j) = P(X_1 = i) \times P(X_2 = j)$$

Resulting distribution:

$$Y = X_1 + X_2 \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 6 \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

2.1 Bernoulli Distribution

2.1.1 Definition

A Bernoulli distribution models binary events; the result is either *failure* (0) or *success* (1). It has a parameter p which represents the probability of success (and $1 - p$ is the probability of failure).

2.1.2 Example

I throw a rigged coin that has probability $\frac{1}{3}$ of being heads and $\frac{2}{3}$ of being tails; I am trying to get heads.

This can be modeled as $X \sim \text{Bernoulli}\left(\frac{1}{3}\right)$. The pmf is

$$P(X = x) = \begin{cases} \frac{1}{3}, & x = 1, \\ \frac{2}{3}, & x = 0. \end{cases}$$

(Represented as a small table)

$$X \sim \begin{pmatrix} 1 & 0 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

2.1.3 Mean and Variance

Mean: E[X] = p
Variance: Var[X] = p(1-p)

Intuition: On average, if we make a lot of experiments, the probability of success will converge to p.

2.2 Binomial Distribution

2.2.1 Definition

This models multiple independent Bernoulli trials. It has two parameters: n — number of trials, and p — success probability per trial.

2.2.2 Example

With the same (rigged) coin, what is the probability that I get 3 heads in 5 throws?

Let X be the number of heads in $n = 5$ independent throws. Each throw is Bernoulli($p = \frac{1}{3}$).

a) How many sequences of 5 throws contain 3 heads?

The number of sequences of length n with exactly k heads is the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

For this example ($n = 5,\ k = 3$):

$$\binom{5}{3} = \frac{5!}{3!2!} = 10.$$

(general) Number of sequences with exactly k successes in n trials: $\binom{n}{k}$.

b) The actual problem: probability of exactly 3 heads in 5 throws

$$P(X = 3) = \binom{5}{3} \left(\tfrac{1}{3}\right)^3 \left(\tfrac{2}{3}\right)^2 = 10 \cdot \frac{1}{27} \cdot \frac{4}{9} = \frac{40}{243} \approx 0.1646.$$

Explanation:
$(\frac{1}{3})^3$ means we want 3 heads, $(\frac{2}{3})^2$ means we also want 2 tails. This results in exactly 3 heads.

Binomial distribution (general and current example).

General:

$$X \sim \text{Binomial}(n, p), \qquad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Current example: $X \sim \text{Binomial}(5, \frac{1}{3})$. The PMF table is

k	0	1	2	3	4	5
$\binom{5}{k}$	1	5	10	10	5	1
$P(X = k)$	$\frac{32}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{40}{243}$	$\frac{10}{243}$	$\frac{1}{243}$

2.2.3 Right Tail Probability

What is the probability that I get at least 3 heads in 5 throws?

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) = \frac{\binom{5}{3}(1/3)^3(2/3)^2 + \binom{5}{4}(1/3)^4(2/3)^1 + \binom{5}{5}(1/3)^5}{1}.$$

Compute the numerators (common denominator 243):

$$P(X \geq 3) = \frac{40 + 10 + 1}{243} = \frac{51}{243} = \frac{17}{81} \approx 0.2099.$$

(general):

$$P(X \geq k_0) = \sum_{k=k_0}^{n} \binom{n}{k} p^k (1-p)^{n-k}.$$

2.2.4 Mean and Variance

For a Binomial distribution $X \sim \text{Binomial}(n, p)$:

$$E[X] = np, \qquad Var(X) = np(1-p)$$

For our example ($n = 5, p = \frac{1}{3}$):

$$E[X] = 5 \times \tfrac{1}{3} = \tfrac{5}{3} \approx 1.6667$$

$$Var(X) = 5 \times \tfrac{1}{3} \times \tfrac{2}{3} = \tfrac{10}{9} \approx 1.1111$$

$$\sigma_X = \sqrt{Var(X)} = \sqrt{\tfrac{10}{9}} \approx 1.054$$

Intuition for Mean and Variance: Remember that Bin(n, p) is just n consecutive independent Bernoulli(p). That means Bin(n, p) = $\sum_{i=1}^{n} Bernoulli(p) = n \times Bernoulli(p)$. Now $E[Bin(n, p)] = E[n \times Bernoulli(p)] = n \times E[Bernoulli(p)] = n \times p$.
Variance is an exercise for you

2.2.5 Additivity Property

If $X_1 \sim \text{Binomial}(n_1, p)$ and $X_2 \sim \text{Binomial}(n_2, p)$ are **independent** random variables with the **same success probability** p, then:

$$X_1 + X_2 \sim \text{Binomial}(n_1 + n_2, p).$$

Example: Suppose you flip a coin with $p = 0.4$ success probability 10 times, then flip the same coin 15 more times. If X_1 is the number of successes in the first 10 flips and X_2 is the number in the next 15 flips, then:

$$X_1 \sim \text{Binomial}(10, 0.4), \quad X_2 \sim \text{Binomial}(15, 0.4),$$

and the total number of successes is:

$$X_1 + X_2 \sim \text{Binomial}(25, 0.4).$$

Note: This property only holds when both binomial distributions have the **same** parameter p. If the success probabilities differ, the sum is not binomial.

2.3 Geometric Distribution

2.3.1 Definition

Geom(p) models the number of independent Bernoulli(p) probes until a success.

2.3.2 Example

I throw the same coin from earlier and I stop when I get heads, what is the probability of having to throw it 5 times?

This means the first 4 throws were tails and the 5th was heads.

$$P(X = 5) = \left(\frac{2}{3}\right)^4 \times \frac{1}{3}$$

Distribution representation from 1 to 5:

x	$P(X = x)$
1	$\frac{1}{3}$
2	$\frac{2}{9}$
3	$\frac{4}{27}$
4	$\frac{8}{81}$
5	$\frac{16}{243}$

General formula:

$$P(X = k) = (1 - p)^{k-1}\, p$$

where $p = \frac{1}{3}$ represents the probability of success (getting heads).

2.3.3 Left Tail Probability

What is the probability of having to throw it less than 3 times?

$$P(X < 3) = P(X = 1) + P(X = 2) = \frac{1}{3} + \frac{2}{9} = \frac{5}{9} \approx 0.5556$$

General:

$$P(X < k_0) = \sum_{i=1}^{k_0-1} (1-p)^{i-1} p = 1 - (1-p)^{k_0-1}$$

2.3.4 Mean and Variance

For a Geometric distribution $X \sim \text{Geom}(p)$:

$$E[X] = \frac{1}{p}, \qquad Var(X) = \frac{1-p}{p^2}.$$

For our example ($p = \frac{1}{3}$):

$$E[X] = \frac{1}{1/3} = 3$$

$$Var(X) = \frac{1 - 1/3}{(1/3)^2} = \frac{2/3}{1/9} = 6$$

$$\sigma_X = \sqrt{Var(X)} = \sqrt{6} \approx 2.449$$

Intuition for mean and variance:
On average, we expect to throw the coin **3 times** until we get heads, with moderate variability ($\sigma_X \approx 2.45$) due to the chance of long sequences of tails.

2.4 Hypergeometric Distribution

Let Z be a random variable that gives the number of black balls that are drawn when a sample of m balls is drawn (without replacement) from a lot of n balls having n_1 black balls and n_2 white balls ($n_1 + n_2 = n$).

The probability mass function of the random variable Z is

$$p_Z(k) = \frac{\binom{n_1}{k}\binom{n-n_1}{m-k}}{\binom{n}{m}}, \quad k = 0, 1, \ldots, \min(n_1, m). \tag{1}$$

We say that Z has a hypergeometric distribution.

2.5 Uniform Distribution

Uniform(n) creates a distribution where every number $v \in \{1, 2, ...n\}$ has the same probability, $\frac{1}{n}$. Simplest example is the distribution of a dice roll which can be modeled with Uniform(6).

Mean: E[X] $= \frac{n+1}{2}$
Variance: $\frac{n^2-1}{12}$

2.6 Poisson Distribution

2.6.1 Definition

A random variable X is **Poisson distributed** with parameter $\lambda > 0$, denoted $X \sim \mathrm{Pois}(\lambda)$ or $X \sim P(\lambda)$, if:

$$X \in \{0, 1, 2, \ldots\} = \mathbb{N}$$

and its probability mass function is given by:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}$$

2.6.2 Interpretation and Usage

The Poisson distribution models the **number of occurrences** of an event of interest in a context where:

- The experiment is repeated a large number of times
- The probability of the event occurring is small

Examples of applications:

- The number of customer arrivals at a counter within a specific time interval
- The number of spam emails received in one minute
- The number of system failures in a given period
- The number of phone calls received during a certain period

2.6.3 Poisson Approximation of the Binomial Distribution

The Poisson distribution arises as the **limit of the binomial distribution** under the following conditions:

Let $X_n \sim \mathrm{Bin}(n, p_n)$ with $n \to \infty$, $p_n \to 0$ and $np_n \to \lambda > 0$.

Then:

$$\lim_{n \to \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Proof of the Limit:

Let $X_n \sim \mathrm{Bin}(n, p_n)$ and let $\lambda = np_n$, which implies $p_n = \frac{\lambda}{n}$. The binomial probability mass function is:

$$P(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k}$$

Substituting $p_n = \frac{\lambda}{n}$:

$$P(X_n = k) = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Rearranging the terms:

$$P(X_n = k) = \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}$$

We analyze the limit of each factor as $n \to \infty$:

1. **Factor 1:** $\lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} = \lim_{n\to\infty} \left(1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right) = 1$
2. **Factor 2:** $\lim_{n\to\infty} \frac{\lambda^k}{k!} = \frac{\lambda^k}{k!}$ (Does not depend on n)
3. **Factor 3:** $\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$ (Definition of the exponential limit)
4. **Factor 4:** $\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = (1-0)^{-k} = 1$

Multiplying the limits of the factors, we obtain:

$$\lim_{n\to\infty} P(X_n = k) = 1 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1 = e^{-\lambda}\frac{\lambda^k}{k!}$$

Practical Rule: The Poisson approximation is useful when:

- n is large (usually $n \geq 20$)
- p is small (usually $p \leq 0.05$)
- The product $\lambda = np$ remains moderate (usually $\lambda \leq 10$)

In this case, we can approximate $\text{Bin}(n, p) \approx \text{Pois}(np)$.
Example: If $X \sim \text{Bin}(100, 0.02)$, then $X \approx \text{Pois}(2)$.

2.6.4 Properties

For $X \sim \text{Pois}(\lambda)$:

- **Mean (Expected Value):** $E[X] = \lambda$
- **Variance:** $\text{Var}(X) = \lambda$
- **Moment Generating Function:** $M_X(t) = e^{\lambda(e^t - 1)}$

Additivity Property: If $X_1 \sim \text{Pois}(\lambda_1)$ and $X_2 \sim \text{Pois}(\lambda_2)$ are independent, then:

$$X_1 + X_2 \sim \text{Pois}(\lambda_1 + \lambda_2)$$

3 Expected Value and Variance

3.1 Graphical interpretation of variance for a discrete random variable

The variance measures how much the values of a discrete random variable differ from their mean. When the probabilities are concentrated around the mean, the variance is small. When the probabilities are spread farther from the mean, the variance becomes large.

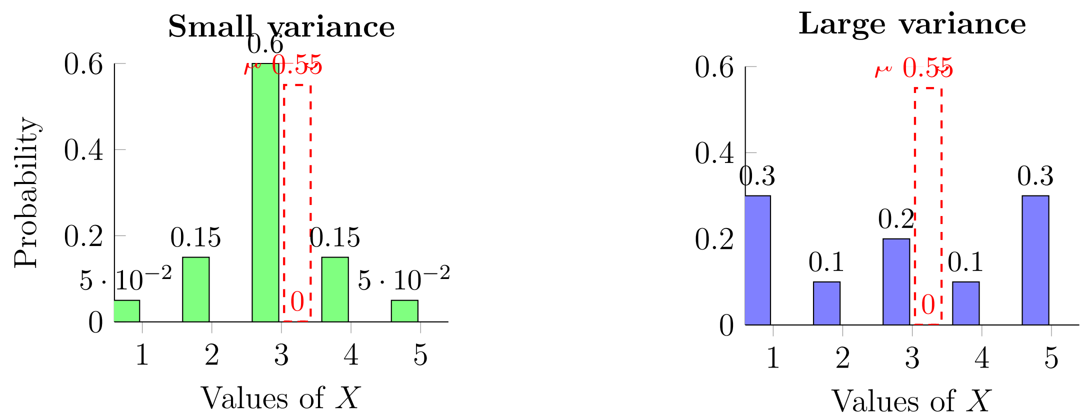


Figure 1: Comparison between a discrete distribution with small variance (left) and large variance (right). When most probability mass is close to the mean, the variance is small; when it is spread far away, the variance increases.

3.2 Visual illustration of variance using discrete points

The concept of variance can also be visualized by representing data as individual points on a number line. When the points are close to the mean, the variance is small. When they are far from the mean, the variance is large.

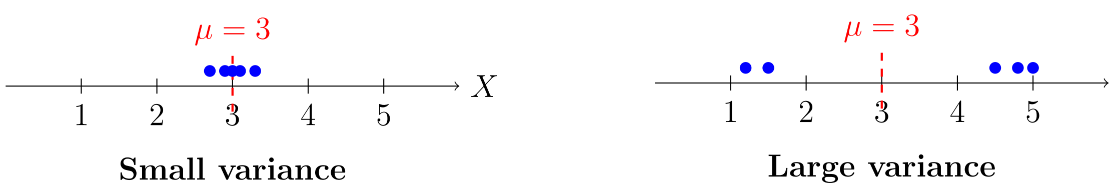


Figure 2: Illustration of variance using discrete points. When data points are close to the mean (left), variance is small. When they are far from the mean (right), variance is large.

3.3 Properties of the Expectation

If $c \in \mathbb{R}$ is a constant and X and Y are random variables defined on the same probability space $(\Omega, \mathcal{K}, P)$, then:

$$E(c) = c, \qquad E(cX) = cE(X), \qquad E(X+Y) = E(X) + E(Y),$$
$$E(X) \leq E(Y), \text{ if } X \leq Y \text{ (i.e. } X(\omega) \leq Y(\omega), \forall \omega \in \Omega).$$

3.4 Definitions and Moments

Definition 2.8. Let $n \in \mathbb{N}$. The real number

$$\beta_n := E(X^n) = \begin{cases} \sum_i x_i^n p_X(x_i), & \text{for discrete } X, \\ \int_{-\infty}^{\infty} x^n f_X(x)\,dx, & \text{for continuous } X. \end{cases}$$

If it exists, it is called the *moment of order n* of X.

The *central moment of order n* of X is

$$\mu_n = E\big[(X-m)^n\big] = \begin{cases} \sum_i (x_i - m)^n p_X(x_i), & \text{for discrete } X, \\ \int_{-\infty}^{\infty} (x-m)^n f_X(x)\,dx, & \text{for continuous } X. \end{cases}$$

Definition 2.12. Let X be a random variable. The *variance* (or *dispersion*) of X is the second central moment of X, denoted by σ_X^2, or simply σ^2, or $\mathrm{Var}(X)$.

Large values of σ_X^2 indicate a wide spread of the values of X around the mean. Conversely, small values of σ_X^2 indicate a concentration of the values of X near the mean. In the extreme case where $\sigma_X^2 = 0$, we have $X = m$ with probability 1 (i.e., the entire distribution mass is concentrated at the mean).

Proposition. The relationship between variance and the moments of X is

$$\sigma^2 = \beta_2 - m^2.$$

Proof.

$$\sigma^2 = E\big[(X-m)^2\big] = E[X^2 - 2mX + m^2] = E[X^2] - 2mE[X] + m^2 = \beta_2 - m^2.$$

Other properties of the variance:

$$\mathrm{Var}(X) \geq 0, \qquad \mathrm{Var}(c) = 0, \qquad \mathrm{Var}(X+c) = \mathrm{Var}(X),\ \forall c \in \mathbb{R}, \qquad \mathrm{Var}(cX) = c^2\mathrm{Var}(X),\ \forall c \in \mathbb{R}.$$

If X is a random variable with mean m, its *standard deviation* is defined as

$$\sigma_X = \sqrt{E\big[(X-m)^2\big]}.$$

An advantage of using σ_X instead of σ_X^2 is that σ_X has the same measurement unit as the mean, which allows direct comparison on the same scale and provides a measure of the degree of dispersion.

A dimensionless number (without measurement units) that characterizes the relative dispersion with respect to the mean and allows comparison between random variables measured in different units is the *coefficient of variation*, defined as

$$v_X = \frac{\sigma_X}{m_X}.$$

3.5 Logarithmic (series) distribution

Consider the discrete distribution (also called the logarithmic or series distribution)

$$P(X = k) = \frac{(1-p)^k}{-k \ln p}, \qquad k = 1, 2, \ldots, \qquad 0 < p < 1.$$

The normalization uses the power series identity

$$\sum_{k=1}^{\infty} \frac{(1-p)^k}{k} = -\ln p,$$

so the pmf sums to 1.

3.6 Expected value

Compute

$$E[X] = \sum_{k=1}^{\infty} k\, P(X = k) = \frac{1}{-\ln p} \sum_{k=1}^{\infty} k \frac{(1-p)^k}{k} = \frac{1}{-\ln p} \sum_{k=1}^{\infty} (1-p)^k.$$

Let $r = 1 - p$ (so $0 < r < 1$). Then

$$\sum_{k=1}^{\infty} r^k = \frac{r}{1-r} = \frac{1-p}{p}.$$

Hence

$$\boxed{E[X] = \frac{1-p}{-p \ln p}}.$$

3.7 Second moment and variance

First,

$$E[X^2] = \sum_{k=1}^{\infty} k^2 P(X = k) = \frac{1}{-\ln p} \sum_{k=1}^{\infty} k^2 \frac{(1-p)^k}{k} = \frac{1}{-\ln p} \sum_{k=1}^{\infty} k(1-p)^k.$$

Use the identity (valid for $|r| < 1$)

$$\sum_{k=1}^{\infty} k r^k = \frac{r}{(1-r)^2}.$$

With $r = 1 - p$ this gives

$$\sum_{k=1}^{\infty} k(1-p)^k = \frac{1-p}{p^2}.$$

Therefore

$$E[X^2] = \frac{1-p}{-p^2 \ln p}.$$

Thus the variance is

$$\mathrm{Var}(X) = E[X^2] - \big(E[X]\big)^2 = \frac{1-p}{-p^2 \ln p} - \left(\frac{1-p}{-p \ln p}\right)^2.$$

A convenient algebraic simplification is

$$\boxed{\mathrm{Var}(X) = \frac{(1-p)\big(-\ln p - (1-p)\big)}{p^2(-\ln p)^2}}.$$

3.8 Worked exercise (numerical check)

Take $p = 0.6$. Then $1 - p = 0.4$ and $-\ln p \approx 0.5108256$.

$$E[X] = \frac{1-p}{-p \ln p} = \frac{0.4}{0.6 \times 0.5108256} \approx 1.305.$$

$$\mathrm{Var}(X) = \frac{(1-p)\big(-\ln p - (1-p)\big)}{p^2(-\ln p)^2} \approx \frac{0.4(0.5108256 - 0.4)}{0.36 \times 0.5108256^2} \approx 0.472.$$

These numerical values are useful to sanity-check an implementation or a simulation.