Normal random variables

Normal or Gaussian random variables are some of the most important examples of continuous random variables. They arise naturally in *the central limit theorem.*

Definition

A continuous random variable X is said to have Gaussian or normal distribution with parameters μ and σ if the probability density function of X is given by

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

A random variable with normal distribution with parameters $\mu = 0$ and $\sigma = 1$ is called a *standard normal distribution* or a standard Gaussian.

Standard normal random variable

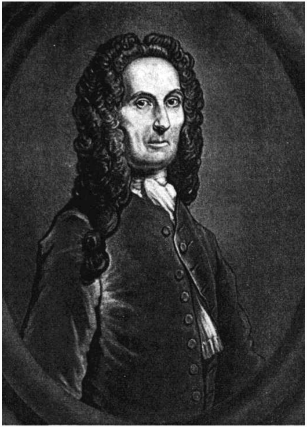
The density function of the standard normal random variable is given by

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

Some basic facts about the standard normal distribution:

- There is no closed formula for $F_X(t)$.
- The distribution function of X is often denoted by $\Phi(t)$, that is $\Phi(t) = \mathbb{P}[X \le t]$, where X has standard normal distribution.
- $\Phi(0) = \frac{1}{2}$.

Who wrote this down and why should we care?



Abraham de Moivre (1667-1754).
Wrote *Doctrine of Chance*
Proved $n! \sim c\sqrt{2n}(n/e)^n$.
Proved the special case of the Central Limit theorem for Bernoulli distributions.

Standard normal random variables

Carl Friedrich Gauss (1777-1855)

Normal distribution of errors in astronomy.

Foretaste of De Moivre's Central Limit Theorem

Question 1: You are throwing a fair coin 100 times.

1. how likely it it for the number of Heads to be between 40 and 50?
2. how likely it is for the number of Heads to be more than 60.

Question 1: You are throwing a fair coin 10000 times.

1. how likely it it for the number of Heads to be between 400 and 500?
2. how likely it is for the number of Heads to be more than 600?

Standardization

Suppose X is a normal random variable with parameters (μ, σ). Then

$$Z = \frac{X - \mu}{\sigma}$$

is a normal random variable with parameters $\mu = 0$ and $\sigma = 1$, that is, a standard normal random variable. The process of going from X to Z is called *standardization.*

Example

Suppose X has normal distribution with parameters $\mu = 3$ and $\sigma = 2$. What is the value of $\mathbb{P}[1 \leq X \leq 7]$?

The random variable $Z = \frac{X-3}{2}$ *is a standard normal random variable. Hence*

$$\begin{aligned} \mathbb{P}[1 \leq X \leq 7] &= \mathbb{P}[1-3 \leq X-3 \leq 7-3] = \mathbb{P}\left[\frac{1-3}{2} \leq \frac{X-3}{2} \leq \frac{7-3}{2}\right] \\ &= \mathbb{P}[-1 \leq Z \leq 2] = \Phi(2) - \Phi(1). \end{aligned} \tag{1}$$

Definition

Let X be a continuous random variable with the density function $f_X(t)$. The expected value of the mean of X is defined by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Example: expected value for uniform random variables

Example

Let X have uniform distribution over the interval $[a, b]$. Then, we have

$$\mathbb{E}\left[X\right] = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}.$$

Note that the answer is consistent with the idea that the expectation corresponds to the center of mass.

Example

Example

Let X be a continuous random variable whose density function is given by

$$f_X(x) = \begin{cases} k(1-x^2) & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of k and Compute $\mathbb{E}[X]$.

$$\int_{-\infty}^{\infty} f_X(x)\ dx = \int_0^1 k(1-x^2)\ dx = 1.$$

$$k(x - \frac{x^3}{3})\Big|_0^1 = \frac{2}{3}k = 1 \Rightarrow k = \frac{3}{2}.$$

$$\mathbb{E}[X] = \int_0^1 x \cdot \frac{3}{2}x(1-x^2)\ dx = \frac{3}{8}.$$

Computing $\mathbb{E}[h(X)]$

Sometimes it is useful to compute the expected value of a *function* of X. The following theorem can be useful:

Theorem

Let X be a continuous random variable with the density function $f_X(t)$. For any continuous function $h(t)$, we have

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x)\, dx.$$

Note that this circumvents computation of the density function for $h(X)$.

Example

Example

Suppose X has the density function given by

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $\mathbb{E}\left[X^2+1\right]$ and

$$\mathbb{E}\left[X^2+1\right] = \int_{-\infty}^{\infty} h(x) f_X(x)\, dx = \int_0^1 (x^2+1)(2x)\, dx = \frac{3}{2}.$$

Example

Example

Suppose X has the distribution function given by

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^2. & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Find $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

$$f_X(x) = \begin{cases} 0 & \text{if } x \leq 0, x > 1 \\ 2x & \text{if } 0 \leq x \leq 1 \end{cases}$$

$$\mathbb{E}[X] = \int_0^1 {\color{red}x} 2x dx = 2\frac{x^3}{3}\Big|_0^1 = \frac{2}{3}, \mathbb{E}\left[X^2\right] = \int_0^1 {\color{red}x^2} 2x dx = 2\frac{x^4}{4}\Big|_0^1 = \frac{1}{2}.$$

$$\mathrm{Var}[X] = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}.$$

Properties of expectation

Theorem

Let X, Y be random variables and c a constant. We have

- *Linearity:* $\mathbb{E}[cX + Y] = c\mathbb{E}[X] + \mathbb{E}[Y]$.
- *Comparison: if* $X \leq Y$ *with probability 1, then* $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Exploiting linearity

Example

n letters are randomly placed into n envelopes. What is the expected number of the letters placed into the right envelope?

X: the number of letters places into the right envelope.

$$X = X_1 + X_2 + \cdots + X_n,$$

X_i: Bernoulli random variables that takes value 1 when the letter i is places in the envelope i.

$$\mathbb{E}\left[X_i\right] = \mathbb{P}\left[X_i = 1\right] = \frac{1}{n},$$

$$\mathbb{E}\left[X\right] = n \cdot 1/n = 1.$$

Example

n people board an elevator on the ground floor of a building with k floors. Each leaves the elevators at one of the floors 1 to k, each chosen randomly and independently from others. Find the expected number of the floors at which the elevator stops.

X_j be the Bernoulli random variable which takes value 1 exactly when the elevator stops at floor j.

$$\mathbb{P}[X_j = 1] = 1 - \left(\frac{k-1}{k}\right)^n.$$

$$X = X_1 + \cdots + X_n.$$

$$\mathbb{E}[X] = \sum_{j=1}^{k} \mathbb{E}[X_j] = k - k\left(\frac{k-1}{k}\right)^n.$$

Large deviation inequalities

Motivating example:

Consider a biased coin that lands heads with probability $\frac{1}{10}$. This coin is flipped 200 times. How large can be the probability that the coin lands heads at least 120 times.

$$\mathbb{P}[A] = \binom{200}{120}(1/10)^{120}(9/10)^{80} + \binom{200}{121}(1/10)^{121}(9/10)^{79} + \cdots + \binom{200}{200}(1/10)^{200}.$$

X: number of heads. Then X has Binomial distribution with $n = 200$ and $p = 1/10$. So,

$$\mathbb{E}[X] = 200 \cdot \frac{1}{10} = 20.$$

This expression is hard to estimate. Is there a way to get an upper bound for this probability?

Large deviation inequalities

Suppose X is a random variable with $\mathbb{E}[X] = \mu$. Can we say something about

$$\mathbb{P}[|X - \mu| > t]$$

The answer is in general no.

- Suppose X is discrete taking values $x_1, \ldots, x_n$ with probabilities $p_1, \ldots, p_n$.
- Suppose $\mathbb{E}[X] = \mu$.

$$\mathrm{Var}[X] = \sum_{i=1}^{n} (x_i - \mu)^2 p_i \geq \sum_{|x_i - \mu| \geq t} (x_i - \mu)^2 p_i \geq t^2 \mathbb{P}[|X - \mu| \geq t].$$

Theorem

For every random variable X with $\mathbb{E}[X] = \mu$ and evert $t > 0$ we have

$$\mathbb{P}[|X - \mu| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}.$$

Back to the example

In the previous example $\mu = 20$.

$$\mathrm{Var}\left[X\right] = np(1-p) = 200 \cdot \frac{1}{10} \cdot \frac{9}{10} = 18.$$

$$\mathbb{P}\left[X \geq 120\right] = \mathbb{P}\left[X - 20 \geq 100\right] = \mathbb{P}\left[|X - 20| \geq 100\right] \leq \frac{18}{100^2} = 0.0018.$$

Joint distribution of random variables

Motivating question Suppose X and Y are two discrete random variables. Knowing $\mathbb{P}[X = x]$ and $\mathbb{P}[Y = y]$, can we say something about

Exploiting linearity

Example

n letters are randomly placed into n envelopes. What is the expected number of the letters placed into the right envelope?

X: the number of letters places into the right envelope.

$$X = X_1 + X_2 + \cdots + X_n,$$

X_i: Bernoulli random variables that takes value 1 when the letter i is places in the envelope i.

$$\mathbb{E}[X_i] = \mathbb{P}[X_i = 1] = \frac{1}{n},$$

$$\mathbb{E}[X] = n \cdot 1/n = 1.$$

Example

n people board an elevator on the ground floor of a building with k floors. Each leaves the elevators at one of the floors 1 to k, each chosen randomly and independently from others. Find the expected number of the floors at which the elevator stops.

X_j be the Bernoulli random variable which takes value 1 exactly when the elevator stops at floor j.

$$\mathbb{P}[X_j = 1] = 1 - \left(\frac{k-1}{k}\right)^n.$$

$$X = X_1 + \cdots + X_n.$$

$$\mathbb{E}[X] = \sum_{j=1}^{k} \mathbb{E}[X_j] = k - k\left(\frac{k-1}{k}\right)^n.$$

Large deviation inequalities

Motivating example:

Consider a biased coin that lands heads with probability $\frac{1}{10}$. This coin is flipped 200 times. How large can be the probability that the coin lands heads at least 120 times.

$$\mathbb{P}\left[A\right] = \binom{200}{120}(1/10)^{120}(9/10)^{80} + \binom{200}{121}(1/10)^{121}(9/10)^{79} + \cdots + \binom{200}{200}(1/10)^{200}.$$

X: number of heads. Then X has Binomial distribution with $n = 200$ and $p = 1/10$. So,

$$\mathbb{E}\left[X\right] = 200 \cdot \frac{1}{10} = 20.$$

This expression is hard to estimate. Is there a way to get an upper bound for this probability?

Large deviation inequalities

Suppose X is a random variable with $\mathbb{E}[X] = \mu$. Can we say something about

$$\mathbb{P}[|X - \mu| > t]$$

The answer is in general no.

- Suppose X is discrete taking values $x_1, \ldots, x_n$ with probabilities $p_1, \ldots, p_n$.
- Suppose $\mathbb{E}[X] = \mu$.

$$\mathrm{Var}[X] = \sum_{i=1}^{n} (x_i - \mu)^2 p_i \geq \sum_{|x_i - \mu| \geq t} (x_i - \mu)^2 p_i \geq t^2 \mathbb{P}[|X - \mu| \geq t].$$

Theorem

For every random variable X with $\mathbb{E}[X] = \mu$ and evert $t > 0$ we have

$$\mathbb{P}[|X - \mu| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}.$$

Back to the example

In the previous example $\mu = 20$.

$$\mathrm{Var}\left[X\right] = np(1-p) = 200 \cdot \frac{1}{10} \cdot \frac{9}{10} = 18.$$

$$\mathbb{P}\left[X \geq 120\right] = \mathbb{P}\left[X - 20 \geq 100\right] = \mathbb{P}\left[|X - 20| \geq 100\right] \leq \frac{18}{100^2} = 0.0018.$$

Joint distribution of random variables

Motivating question Suppose X and Y are two discrete random variables. Knowing

$$\mathbb{P}[X = x], \mathbb{P}[Y = y]$$

can we say something about

$$\mathbb{P}[X = x, Y = y]?$$

Example

Two coins have each been thrown 2 times. Denote

$$\begin{aligned} X &= \text{number of heads for the first coin} \\ Y &= \text{number of heads for the second coin} \\ Z &= \text{number of tails for the first coin} \\ T &= \text{number of tails for the second coin} \end{aligned} \tag{2}$$

What is the distribution of X, Y, Z, T?

Joint distribution of random variables

The distribution for each one of X, Y, Z, T is given by the following table:

x	0	1	2
$X = x$	1/4	1/2	1/4

For values $i, j = 0, 1, 2$, Find the value

$$\mathbb{P}[X = i, Y = j]$$

$$\mathbb{P}[X = i, Z = j]$$

	Y=0	Y=1	Y=2
X=0	1/16	1/8	1/16
X=1	1/8	1/4	1/8
X=2	1/16	1/8	1/16

	Z=0	Z=1	Z=2
X=0	0	0	1/4
X=1	0	1/2	0
X=2	1/4	0	0

Definition

Suppose X and Y are two discrete random variables. Then the joint probability mass function of X and Y is the function $p_{X,Y}(x, y)$ defined by

$$p_{X,Y}(x, y) = \mathbb{P}[X = x, Y = y].$$

One way of presenting the joint probability mass function of X and Y is by using a table:

Example

Suppose X and Y are discrete random variables with the joint probability mass function given by the following table:

	Z=0	Z=1
X=0	1/12	1/12
X=1	5/12	2/12
X=2	1/12	2/12

1. What are the probability mass functions of X and Y?
2. What is the probability mass function of $X + Y$?

X takes values $0, 1, 2$.

	Y=0	Y=1
X=0	1/12	1/12
X=1	5/12	2/12
X=2	1/12	2/12

x	0	1	2
p_X	2/12	7/12	3/12

	Y=0	Y=1
X=0	1/12	1/12

x	0	1

Example

Suppose X and Y are randomly chosen from the set $1, 2, 3, 4$ sich that each pair has the same probability $1/16$ of being chosen. Set $U = \max(X, Y)$ and $V = \min(X, Y)$. Find the joint probability mass function of U and V.

	V=1	V=2	V=3	V=4
U=1	1/16	0	0	0
U=2	2/16	1/16	0	0
U=3	2/16	2/16	1/16	0
U=4	2/16	2/16	2/16	1/16

Independent random variables

Definition

We say that two random variables X and Y are independent when the joint probability mass function of X and Y is given by

$$p_{X,Y}(x,y) = p_X(x)p_Y(y).$$

Equivalently

$$\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x]\,\mathbb{P}[Y = y].$$

Remark: Two random variables X and Y are independent when every event defined in terms of X is independent from any event defined in terms of Y.

Remark: We say that $X_1, X_2, \ldots, X_n$ are independent if

$$\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n] = \mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n].$$

Example

Suppose U and V are random variables from the previous example. Are they independent?

	V=1	V=2	V=3	V=4
U=1	1/16	0	0	0
U=2	2/16	1/16	0	0
U=3	2/16	2/16	1/16	0
U=4	2/16	2/16	2/16	1/16

	V=1	V=2	V=4	V=4	
U=0	1/16	0	0	0	1/16
U=1	2/16	1/16	0	0	3/16
U=3	2/16	2/16	1/16	0	5/16
U=4	2/16	2/16	2/16	1/16	7/16
	7/16	5/16	3/16	1/16	

The Central Limit Theorem

Theorem

(The Central Limit Theorem) Suppose $X_1, X_2, \ldots$ are independent random variables with the same distribution. Suppose that they all have expected value equal μ and variance σ^2. Set

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Then as $n \to \infty$, the distribution of Z_n converges to that of a standard normal random variable.

Note that

$$\mathbb{E}[Z_n] = 0, \qquad \mathrm{Var}[Z_n] = 1.$$

Remark: Often we simply assume that for n large, the distribution of Z_n can be approximated by the distribution of a standard normal random variable.

Example

A fair coin has been tossed 100 times.

1. how likely it it for the number of Heads to be between 45 and 55?
2. how likely it is for the number of Heads to be more than 47.

Let X_i be the Bernoulli random variable which is equal to 1 when the outcome of the i-th toss is Heads. $\mu = \frac{1}{2}$ and $\sigma = \frac{1}{2}$.

We are interested in the distribution of

$$S = X_1 + \cdots + X_{100}.$$

$$\mathbb{P}\left[45 \leq S \leq 55\right] = \mathbb{P}\left[-1 \leq \frac{S - 100 \times \frac{1}{2}}{\frac{1}{2}\sqrt{100}} \leq 1\right].$$

$$\mathbb{P}\left[45 \leq S \leq 55\right] = \Phi(1) - \Phi(-1) \approx 0.84 - 0.15 = 0.69.$$

$$\mathbb{P}\left[47 \leq S\right] = \mathbb{P}\left[-0.6 \leq \frac{S - 100 \times \frac{1}{2}}{\frac{1}{2}\sqrt{100}}\right] = 1 - \Phi(-0.6) \approx 1 - 0.27 = 0.73.$$

Example modified

A fair coin has been tossed 10000 times. How likely it it for the number of Heads to be between 4500 and 5500?

The probability in question can be approximated by

$$\mathbb{P}\left[4500 \leq S \leq 5500\right] = \mathbb{P}\left[-10 \leq \frac{S - 10000 \times \frac{1}{2}}{\frac{1}{2}\sqrt{10000}} \leq 10\right] \approx 1.$$