Chapter 5B: Normal and Exponential Distributions

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

Summer 2026

Learning Goals

By the end of this lecture, students should be able to:

- define the standard Normal distribution;
- standardize a Normal random variable;
- compute Normal probabilities using the standard Normal CDF;
- use the 68–95–99.7 rule;
- define the Exponential distribution;
- compute Exponential probabilities;
- understand the memoryless property of the Exponential distribution.

Normal Distribution

The Normal Distribution $N(\mu, \sigma^2)$

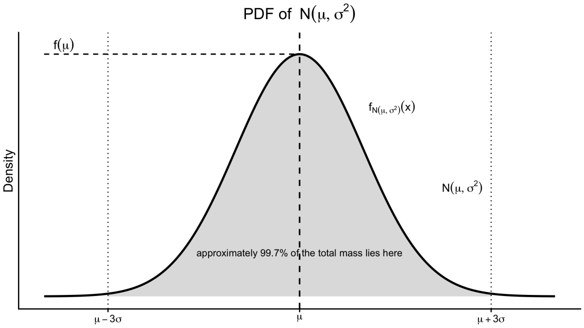
Definition

A continuous random variable X has a Normal distribution with mean μ and variance σ^2 if

$$X \sim N(\mu, \sigma^2),$$

and its density is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}.$$



The parameter μ controls the center, while σ^2 controls the spread. The special case where $\mu = 0$, $\sigma = 1$ is called Standard Normal.

Preview

This is the most fundamental distribution in probability, with so many important fetures. Central Limit Theorem (CLT) will later shows one of its key aspects.

PDF Verification of Normal

$f_X(x) \geq 0$ is straightforward, so it suffices to show

Claim

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

Let

$$I = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx.$$

Then

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-x^2/2}\, dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}\, dy\right) = \int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\, dx\, dy.$$

Using polar coordinates,

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad dx\, dy = r\, dr\, d\theta.$$

Therefore,

$$I^2 = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r\, dr\, d\theta.$$

Let $u = r^2/2$, so $du = r\, dr$. Then

$$I^2 = \int_0^{2\pi}\int_0^{\infty} e^{-u}\, du\, d\theta = \int_0^{2\pi} 1\, d\theta = 2\pi.$$

Thus,

$$I = \sqrt{2\pi}.$$

Mean and Variance of $N(\mu, \sigma^2)$

Proposition

Suppose

$$X \sim N(\mu, \sigma^2),$$

with density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Then,

$$\mathbb{E}(X) = \mu, \qquad \mathrm{Var}(X) = \sigma^2.$$

Proof: By definition,

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

Let

$$z = x - \mu.$$

Then

$$x = z + \mu, \qquad dx = dz.$$

Therefore,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} (z + \mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/(2\sigma^2)}\, dz.$$

Splitting the integral,

$$\mathbb{E}(X) = \underbrace{\int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/(2\sigma^2)}\, dz}_{=0} + \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/(2\sigma^2)}\, dz}_{=1}.$$

Hence,

$$\mathbb{E}(X) = \mu.$$

Variance of $N(\mu, \sigma^2)$

Recall:

$$\mathrm{Var}(X) = \mathbb{E}\big[(X-\mu)^2\big].$$

Using LOTUS,

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx.$$

Let

$$z = x - \mu.$$

Then

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/(2\sigma^2)}\,dz.$$

Now substitute

$$u = \frac{z}{\sigma}, \qquad z = \sigma u, \qquad dz = \sigma\,du.$$

Then

$$\mathrm{Var}(X) = \sigma^2 \int_{-\infty}^{\infty} u^2 \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,du.$$

The remaining integral equals 1, since it is the variance of the standard Normal distribution $N(0,1)$.
Therefore,

$$\mathrm{Var}(X) = \sigma^2.$$

Standard Normal CDF

The standard Normal CDF is denoted by

$$\Phi(z).$$

It is defined by

$$\Phi(z) = \mathbb{P}(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

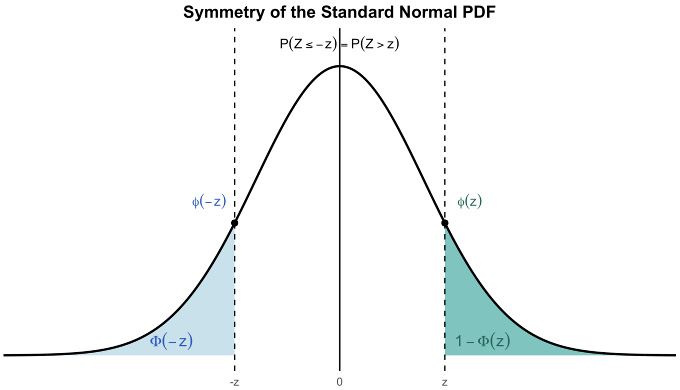
Important

There is no elementary closed-form expression for $\Phi(z)$.

So Normal probabilities are usually computed using tables, calculators, or software. See "normal-table.pdf" in Canvas for the Standard Normal table.

Symmetry of the Standard Normal

The standard Normal PDF is symmetric about 0, i.e., $\varphi(z) = \varphi(-z)$. Therefore, $\mathbb{P}(Z \le -z) = \mathbb{P}(Z \ge z)$. In terms of the CDF, $\Phi(-z) = 1 - \Phi(z)$. This is why some Normal CDF tables skip the negative values and only provide $\phi(z)$ for $z \ge 0$.



Converting $N(0,1)$ to $N(\mu, \sigma^2)$

Proposition

If
$$Z \sim N(0,1),$$
then
$$X = \mu + \sigma Z$$
has the Normal distribution with mean μ and variance σ^2, i.e.,
$$X \sim N(\mu, \sigma^2).$$

Converting $N(\mu, \sigma^2)$ to $N(0, 1)$

Proposition

If

$$X \sim N(\mu, \sigma^2),$$

then

$$\boxed{Z = \frac{X - \mu}{\sigma} \sim N(0, 1).}$$

Standardization converts a general Normal random variable into a standard Normal.

Use

To compute probabilities involving X, convert them into probabilities involving Z, and then use the Normal probability table.

Example: Standardizing

Suppose

$$X \sim N(70, 10^2).$$

Find

$$\mathbb{P}(X < 85).$$

Standardize:

$$\mathbb{P}(X < 85) = \mathbb{P}\left(\frac{X - 70}{10} < \frac{85 - 70}{10}\right).$$

Thus

$$\mathbb{P}(X < 85) = \mathbb{P}(Z < 1.5) = \Phi(1.5).$$

The 68–95–99.7 Rule

If

$$X \sim N(\mu, \sigma^2),$$

then approximately:

$$\mathbb{P}(|X - \mu| < \sigma) \approx 0.68,$$

$$\mathbb{P}(|X - \mu| < 2\sigma) \approx 0.95,$$

$$\mathbb{P}(|X - \mu| < 3\sigma) \approx 0.997.$$

Interpretation

Most Normal observations lie within a few standard deviations of the mean.

Normal Approximation to the Binomial

Motivation: Large Binomial Probabilities

Suppose

$$X \sim \mathrm{Bin}(n, p).$$

For a single value k, we can compute

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

However, probabilities such as

$$\mathbb{P}(X \leq k) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j}$$

may require adding a very large number of terms.

Main idea

When n is sufficiently large, the Binomial distribution is approximately Normal.

Thus, complicated Binomial sums can often be approximated using the standard Normal CDF Φ.

Why Does the Normal Approximation Work?

A Binomial random variable can be written as

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

where

$$X_i = \begin{cases} 1, & \text{if trial } i \text{ is a success,} \\ 0, & \text{if trial } i \text{ is a failure.} \end{cases}$$

The variables $X_1, \ldots, X_n$ are independent Bernoulli random variables with

$$\mathbb{E}(X_i) = p, \qquad \mathrm{Var}(X_i) = p(1-p).$$

Therefore,

$$\mathbb{E}(X) = np, \qquad \mathrm{Var}(X) = np(1-p).$$

As n becomes large, the distribution of the sum

$$X_1 + \cdots + X_n$$

becomes approximately bell-shaped.

Preview

This is a consequence of the Central Limit Theorem, which will be studied later.

Approximation

Suppose

$$X \sim \mathrm{Bin}(n, p).$$

When n is sufficiently large and p is not too close to 0 or 1, we may use the approximation

$$\boxed{X \approx Y, \qquad Y \sim N\big(np, np(1-p)\big).}$$

Thus, the approximating Normal random variable has the same mean and variance as the Binomial random variable:

$$\mu = np, \qquad \sigma^2 = np(1-p), \qquad \sigma = \sqrt{np(1-p)}.$$

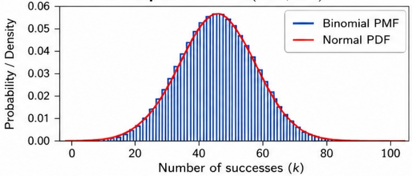
After applying a continuity correction, we standardize using

$$Z = \frac{Y - np}{\sqrt{np(1-p)}}.$$

Normal Approximation to the Binomial Distribution

If $X \sim \text{Bin}(n, p)$ with $np \geq 10$ and $n(1-p) \geq 10$, then $\quad X \approx N(np, np(1-p))$

Example: $X \sim \text{Bin}(100, 0.4)$



Key quantities:

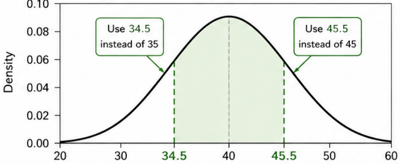
$\mu = np = 40$

$\sigma = \sqrt{np(1-p)} = \sqrt{24} \approx 4.90$

Here, $np = 40 \geq 10$ and $n(1-p) = 60 \geq 10$

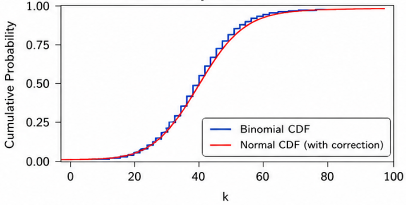
The binomial distribution is well approximated by the normal distribution.

Approximating $P(35 \leq X \leq 45)$



$$\begin{aligned} P(35 \leq X \leq 45) &\approx P(34.5 < Y < 45.5) \\ &= P\left(\frac{34.5-40}{4.90} < Z < \frac{45.5-40}{4.90}\right) \\ &= P(-1.12 < Z < 1.12) \approx 0.739 \end{aligned}$$

CDF Comparison



The normal approximation becomes more accurate as n increases.

When Is the Approximation Appropriate?

A commonly used rule is to require

$$\boxed{np \geq 10 \qquad \text{and} \qquad n(1-p) \geq 10.}$$

These quantities represent the expected numbers of successes and failures.

Interpretation

The approximation works best when we expect sufficiently many successes and sufficiently many failures.

If p is very close to 0, the Binomial distribution is strongly right-skewed.
If p is very close to 1, the Binomial distribution is strongly left-skewed.
In either case, a symmetric Normal curve may provide a poor approximation.

Why Is a Continuity Correction Needed?

A Binomial random variable is discrete:

$$X \in \{0, 1, 2, \dots, n\}.$$

A Normal random variable is continuous:

$$Y \in \mathbb{R}.$$

For example, the event $X = 5$ is represented by the interval centered at 5:

$$4.5 < Y < 5.5.$$

Thus,

$$\mathbb{P}(X = 5) \approx \mathbb{P}(4.5 < Y < 5.5).$$

Continuity correction

To translate a discrete boundary into a continuous boundary, shift the endpoint by 0.5.

Continuity-Correction Rules

Let

$$X \sim \mathrm{Bin}(n, p), \qquad Y \sim N\big(np, np(1-p)\big).$$

Once standardized, the following approximations are used:

Binomial probability	Normal approximation
$\mathbb{P}(X = k)$	$\mathbb{P}(k - 0.5 < Y < k + 0.5)$
$\mathbb{P}(X \leq k)$	$\mathbb{P}(Y < k + 0.5)$
$\mathbb{P}(X < k)$	$\mathbb{P}(Y < k - 0.5)$
$\mathbb{P}(X \geq k)$	$\mathbb{P}(Y > k - 0.5)$
$\mathbb{P}(X > k)$	$\mathbb{P}(Y > k + 0.5)$
$\mathbb{P}(a \leq X \leq b)$	$\mathbb{P}(a - 0.5 < Y < b + 0.5)$

Procedure for the Normal Approximation

Suppose

$$X \sim \text{Bin}(n, p).$$

1. Check if the conditions hold

$$np \geq 10, \qquad n(1-p) \geq 10.$$

2. Compute

$$\mu = np, \qquad \sigma = \sqrt{np(1-p)}.$$

3. Apply the continuity correction.
4. Standardize using

$$Z = \frac{Y - \mu}{\sigma}.$$

5. Use the standard Normal CDF Φ.

Important

The continuity correction must be applied before standardizing.

Example: A Binomial Probability

Suppose

$$X \sim \text{Bin}(100, 0.4).$$

Approximate

$$\mathbb{P}(35 \leq X \leq 45).$$

First,

$$np = 40, \qquad n(1-p) = 60,$$

so the Normal approximation is appropriate.
The mean and standard deviation are

$$\mu = 100(0.4) = 40,$$

and

$$\sigma = \sqrt{100(0.4)(0.6)} = \sqrt{24}.$$

Using the continuity correction,

$$\mathbb{P}(35 \leq X \leq 45) \approx \mathbb{P}(34.5 < Y < 45.5),$$

where

$$Y \sim N(40, 24).$$

Example: A Binomial Probability, Continued

Standardizing gives

$$\mathbb{P}(34.5 < Y < 45.5) = \mathbb{P}\left(\frac{34.5-40}{\sqrt{24}} < Z < \frac{45.5-40}{\sqrt{24}}\right).$$

Therefore,

$$\mathbb{P}(35 \leq X \leq 45) \approx \mathbb{P}(-1.123 < Z < 1.123).$$

Using the standard Normal CDF,

$$\mathbb{P}(-1.123 < Z < 1.123) = \Phi(1.123) - \Phi(-1.123).$$

By symmetry,

$$\Phi(-1.123) = 1 - \Phi(1.123).$$

Thus,

$$\mathbb{P}(35 \leq X \leq 45) \approx 2\Phi(1.123) - 1 \approx 0.739.$$

Example: Stocking a Bestseller

Suppose that the odds are 1 to 5000 in favor of a customer buying a particular fiction bestseller.
If 800 customers enter the store every day, how many copies should the store stock each month so that, with probability greater than 98%, it does not run out?
Assume that a month has 30 days.

Source: Saeed Ghahramani, *Fundamentals of Probability*.

Solution: Stocking a Bestseller

Let

$$X = \text{number of customers who buy the book during the month.}$$

The total number of customers is

$$n = 30(800) = 24000.$$

Therefore,

$$X \sim \text{Bin}\left(24000, \frac{1}{5000+1}\right).$$

The expected number of buyers is

$$np = \frac{24000}{5000} \approx 4.799.$$

However,

$$np < 10.$$

Conclusion

The usual Normal approximation is not appropriate because the expected number of successes is too small.

Solution: Using a Poisson Approximation

Since n is large and p is very small, the Binomial distribution is better approximated by a Poisson distribution:

$$X \approx \text{Pois}(\lambda), \qquad \lambda = np \approx 4.799.$$

If the bookstore stocks m copies, it does not run out when

$$X \leq m.$$

We seek the smallest integer m such that

$$\mathbb{P}(X \leq m) > 0.98.$$

Using the Poisson CDF,

$$\mathbb{P}(X \leq 9) \approx \sum_{k=0}^{9} e^{-4.799} \frac{4.799^k}{k!} \approx 0.975,$$

which is less than 0.98.
However,

$$\mathbb{P}(X \leq 10) \approx \sum_{k=0}^{10} e^{-4.799} \frac{4.u8^k}{k!} \approx 0.990.$$

Therefore, the store should stock

$$\boxed{\text{10 copies.}}$$

Example: Sampling Light Bulbs

Every day, a factory produces 5000 light bulbs, of which 2500 are Type I and 2500 are Type II.
A sample of 40 light bulbs is selected at random without replacement.
What is the approximate probability that the sample contains at least 18 light bulbs of each type?

Source: Saeed Ghahramani, *Fundamentals of Probability*.

Solution: Defining the Random Variable

Let

$$X = \text{number of Type I bulbs in the sample.}$$

Exactly,

$$X \sim \text{Hypergeo}(5000, 2500, 40).$$

Because the sample size is small relative to the population size,

$$\frac{40}{5000} = 0.008,$$

sampling without replacement is approximately equivalent to sampling with replacement.
Thus,

$$X \approx \text{Bin}\left(40, \frac{2500}{5000}\right) = \text{Bin}\left(40, \frac{1}{2}\right).$$

We have

$$np = 20, \qquad n(1-p) = 20,$$

so the Normal approximation is appropriate.

Solution: Translating the Event

If the sample contains X Type I bulbs, then it contains

$$40 - X$$

Type II bulbs.
We require at least 18 bulbs of each type:

$$X \geq 18$$

and

$$40 - X \geq 18.$$

The second inequality gives

$$X \leq 22.$$

Therefore, the desired event is

$$18 \leq X \leq 22.$$

For the Binomial approximation,

$$\mu = np = 20,$$

and

$$\sigma^2 = np(1 - p) = 10.$$

Thus,

$$X \approx Y, \qquad Y \sim N(20, 10).$$

Solution: Applying the Continuity Correction

Using the continuity correction,

$$\mathbb{P}(18 \leq X \leq 22) \approx \mathbb{P}(17.5 < Y < 22.5).$$

Standardizing,

$$\mathbb{P}(17.5 < Y < 22.5) = \mathbb{P}\left(\frac{17.5 - 20}{\sqrt{10}} < Z < \frac{22.5 - 20}{\sqrt{10}}\right).$$

Since

$$\frac{2.5}{\sqrt{10}} \approx 0.791,$$

we obtain

$$\mathbb{P}(18 \leq X \leq 22) \approx \mathbb{P}(-0.791 < Z < 0.791).$$

Therefore,

$$\mathbb{P}(18 \leq X \leq 22) \approx \Phi(0.791) - \Phi(-0.791).$$

Solution: Final Probability

Using symmetry,

$$\Phi(-0.791) = 1 - \Phi(0.791).$$

Hence,

$$\mathbb{P}(-0.791 < Z < 0.791) = 2\Phi(0.791) - 1.$$

Using

$$\Phi(0.791) \approx 0.785,$$

we obtain

$$2(0.785) - 1 \approx 0.570.$$

Therefore,

$$\boxed{\mathbb{P}(\text{at least 18 bulbs of each type}) \approx 0.570.}$$

Interpretation

There is approximately a 57% probability that the sample contains between 18 and 22 Type I bulbs, and hence at least 18 bulbs of each type.

Example: A Company's Lifetime Claim

Suppose that the lifetimes of light bulbs produced by a company are Normal random variables with mean 1000 hours and standard deviation 100 hours.
The company claims that 95% of its light bulbs last at least 900 hours.
Is the company's claim correct?

Source: Saeed Ghahramani, *Fundamentals of Probability*.

Let

$$X = \text{lifetime of a randomly selected light bulb.}$$

Then

$$X \sim N(1000, 100^2).$$

We compute

$$\mathbb{P}(X \geq 900).$$

Standardizing,

$$\mathbb{P}(X \geq 900) = \mathbb{P}\left(Z \geq \frac{900 - 1000}{100}\right).$$

Therefore,

$$\mathbb{P}(X \geq 900) = \mathbb{P}(Z \geq -1).$$

By symmetry,

$$\mathbb{P}(Z \geq -1) = \Phi(1).$$

Thus,

$$\mathbb{P}(X \geq 900) \approx 0.8413.$$

Solution: Evaluating the Claim

The actual probability is

$$\mathbb{P}(X \geq 900) \approx 0.8413.$$

Thus, approximately 84.13% of the bulbs last at least 900 hours.
The company claims that the percentage is 95%.
Since

$$0.8413 < 0.95,$$

the claim is incorrect.

$$\boxed{\text{No, the company's claim is not correct.}}$$

Remark

A lifetime of 900 hours is only one standard deviation below the mean. For a Normal distribution, approximately 84%, not 95%, of the observations lie above this point.

Example: Bulbs from Two Companies

The lifetime of a bulb produced by the first company is Normal with mean 1000 hours and standard deviation 100 hours.
The lifetime of a bulb produced by the second company is Normal with mean 900 hours and standard deviation 150 hours.
Howard buys one bulb from each company.
Assuming the two lifetimes are independent, what is the probability that at least one bulb lasts 980 or more hours?

Source: Saeed Ghahramani, *Fundamentals of Probability*.

Solution: Defining the Lifetimes

Let

$$X = \text{lifetime of the bulb from the first company},$$

and

$$Y = \text{lifetime of the bulb from the second company}.$$

Then

$$X \sim N(1000, 100^2),$$

and

$$Y \sim N(900, 150^2).$$

We want

$$\mathbb{P}(X \geq 980 \text{ or } Y \geq 980).$$

It is easier to use the complementary event:

$$\mathbb{P}(X \geq 980 \text{ or } Y \geq 980) = 1 - \mathbb{P}(X < 980,\ Y < 980).$$

Because X and Y are independent,

$$\mathbb{P}(X < 980,\ Y < 980) = \mathbb{P}(X < 980)\mathbb{P}(Y < 980).$$

Solution: First Company's Bulb

For the first bulb,

$$\mathbb{P}(X < 980) = \mathbb{P}\left(Z < \frac{980 - 1000}{100}\right).$$

Thus,

$$\mathbb{P}(X < 980) = \mathbb{P}(Z < -0.2) = \Phi(-0.2).$$

Using the Normal table,

$$\Phi(-0.2) \approx 0.4207.$$

Therefore,

$$\mathbb{P}(X < 980) \approx 0.4207.$$

Equivalently,

$$\mathbb{P}(X \geq 980) \approx 1 - 0.4207 = 0.5793.$$

Solution: Second Company's Bulb

For the second bulb,

$$\mathbb{P}(Y < 980) = \mathbb{P}\left(Z < \frac{980 - 900}{150}\right).$$

Now,

$$\frac{980 - 900}{150} = \frac{80}{150} \approx 0.533.$$

Therefore,

$$\mathbb{P}(Y < 980) = \Phi(0.533) \approx 0.7031.$$

Equivalently,

$$\mathbb{P}(Y \geq 980) \approx 1 - 0.7031 = 0.2969.$$

Solution: At Least One Bulb

Using independence,

$$\mathbb{P}(X < 980,\ \ Y < 980) = \mathbb{P}(X < 980)\mathbb{P}(Y < 980).$$

Thus,

$$\mathbb{P}(X < 980,\ \ Y < 980) \approx (0.4207)(0.7031) \approx 0.2958.$$

Therefore,

$$\mathbb{P}(X \geq 980 \text{ or } Y \geq 980) \approx 1 - 0.2958.$$

Hence,

$$\boxed{\mathbb{P}(\text{at least one lasts at least 980 hours}) \approx 0.7042.}$$

Normal Approximation: Summary

Suppose

$$X \sim \text{Bin}(n, p).$$

When

$$np \geq 10 \qquad \text{and} \qquad n(1-p) \geq 10,$$

we use

$$X \approx N\big(np, np(1-p)\big).$$

The procedure is:

1. Compute
$$\mu = np, \qquad \sigma = \sqrt{np(1-p)}.$$
2. Apply the continuity correction.
3. Standardize:
$$Z = \frac{Y - \mu}{\sigma}.$$
4. Use Φ to calculate the probability.

Important

If np or $n(1-p)$ is too small, another approximation, such as the Poisson approximation, may be more appropriate.

Exponential Distribution

Summary so far.

The Exponential distribution models waiting times.

Examples:

- waiting time until a phone call;
- waiting time until a radioactive particle decays;
- waiting time until a customer arrives;
- lifetime of certain components.

It is the continuous analog of waiting for the first success.

Definition of Exponential Distribution

Definition

A random variable X has the Exponential distribution with rate $\lambda > 0$ if its PDF is

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0.$$

We write

$$X \sim \text{Exp}(\lambda).$$

For $x \leq 0$,

$$f(x) = 0.$$

Exponential CDF and Survival Function

If

$$X \sim \exp(\lambda),$$

then for $x > 0$,

$$F(x) = \mathbb{P}(X \leq x) = 1 - e^{-\lambda x}.$$

Therefore,

$$\mathbb{P}(X > x) = 1 - F(x) = e^{-\lambda x}.$$

Survival function

$$\boxed{\mathbb{P}(X > x) = e^{-\lambda x}.}$$

Mean and Variance of Exponential

If

$$X \sim \mathrm{Exp}(\lambda),$$

then

$$\mathbb{E}(X) = \frac{1}{\lambda},$$

and

$$\mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Interpretation

A larger rate λ means events occur more frequently, so waiting times become shorter on average.

Proof of the Mean

The density of an exponential random variable is

$$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0.$$

Thus

$$\mathbb{E}(X) = \int_0^\infty x\lambda e^{-\lambda x}\, dx.$$

Use integration by parts.
Let

$$u = x, \qquad dv = \lambda e^{-\lambda x}\, dx.$$

Then

$$du = dx, \qquad v = -e^{-\lambda x}.$$

Therefore,

$$\mathbb{E}(X) = \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\, dx.$$

The boundary term equals 0, so

$$\mathbb{E}(X) = \int_0^\infty e^{-\lambda x}\, dx = \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^\infty.$$

Hence

$$\mathbb{E}(X) = \frac{1}{\lambda}.$$

Proof of the Variance

We first compute

$$\mathbb{E}(X^2).$$

Using the density,

$$\mathbb{E}(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x}\, dx.$$

Applying integration by parts twice gives

$$\mathbb{E}(X^2) = \frac{2}{\lambda^2}.$$

Now use

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Therefore,

$$\mathrm{Var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2.$$

Hence

$$\mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Memoryless Property of $Exp(\lambda)$

Memoryless Property

If
$$X \sim \exp(\lambda),$$
then for $s, t \geq 0$,
$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t).$$

Interpretation:
Given that we have already waited s units of time, the additional waiting time behaves like a fresh Exponential random variable. In other words, how long you have already waited does not affect how much longer you will need to wait

Remark: Exponential distribution is the only memoryless r.v. among all the continuous r.v.s. Also, the only memoryless discrete r.v. is the Geometric r.v. that we studied in Chapter 3.

Proof of Memorylessness

Using conditional probability,

$$\mathbb{P}(X > s + t \mid X > s) = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)}.$$

For an Exponential random variable,

$$\mathbb{P}(X > x) = e^{-\lambda x}.$$

Thus,

$$\frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}.$$

But

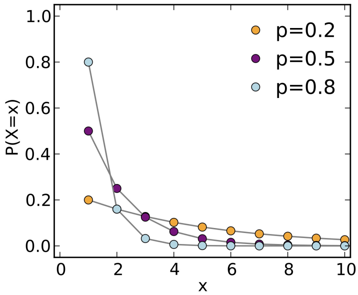
$$e^{-\lambda t} = \mathbb{P}(X > t).$$

Therefore,

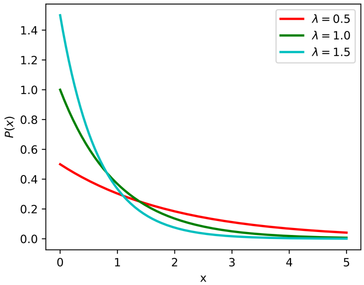
$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t).$$

Both distributions model a **waiting time until the first event**.

Distribution	Time type	Function	Formula
Geom(p)	discrete	PMF	$\mathbb{P}(X = x) = \frac{p}{1-p} e^{\ln(1-p)x}, \quad x \in \mathbb{Z}^+$
$\exp(\lambda)$	continuous	PDF	$f(x) = \lambda e^{-\lambda x}, \quad x \in \mathbb{R}^+$



Geometric PMF: probability at integer times



Exponential PDF: density over continuous time

Summary

- The Normal distribution is bell-shaped and central in statistics.
- Standard Normal:
$$Z \sim N(0, 1).$$
- Standardization:
$$\frac{X - \mu}{\sigma} \sim N(0, 1).$$
- Exponential random variables model waiting times.
- If $X \sim \exp(\lambda)$, then
$$\mathbb{P}(X > x) = e^{-\lambda x}.$$
- The Exponential distribution is memoryless.

Next lecture

Joint distributions and transformations.