The Normal Model

STAT-UB.0001 Statistics for Business Control

Ningshan Zhang

IOMS Department
nzhang@stern.nyu.edu

Jul 24, 2018

* Office Hours: Wed & Fri 10:00 - 11:30 AM, KMC 8-174

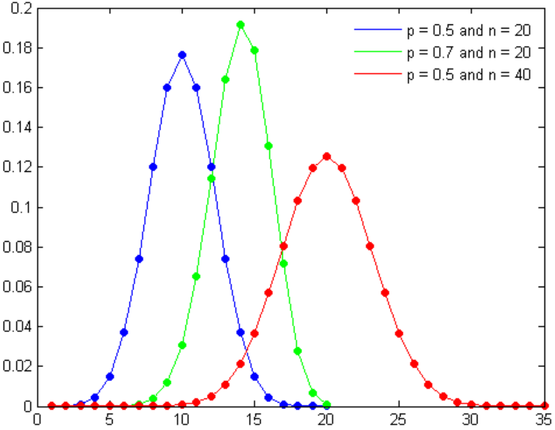
Review: Models for Counts

Two common distributions for counts:

- Binomial distribution, $B(n, p)$
  - X $=$ number of successes of n independent trials, each has probability p of success.
- Poisson distribution, Poisson(λ)
  - X $=$ the number of events that occur in a fixed interval, where events occur with a constant rate, and the events occur independently.

Review: Binomial Distribution

Figure: The PDF of binomial distribution $B(n, p)$.



Review: Poisson Distribution

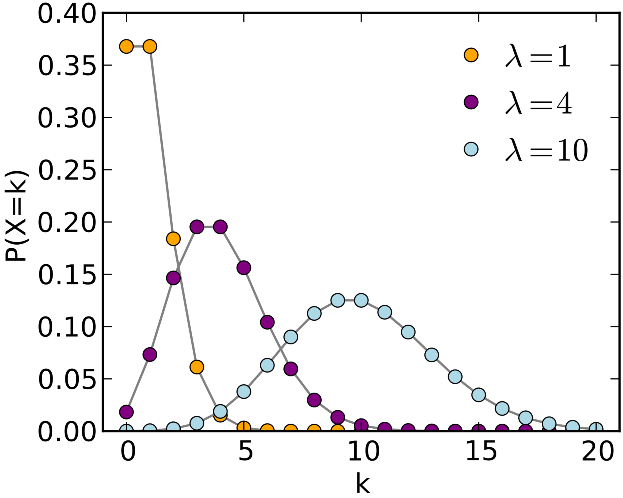
When to use Poisson distribution: counting the number of successes, where

- There is a large number of trials;
- The probability of success for each trial is small;
- The long run average number of successes within a fixed interval is known.

Example: (mostly used) number of arrivals in a time window; number of emails you get in an hour; number of appearances of a word in a document.

Review: Poisson Distribution

Figure: The PDF of Poisson distribution Poisson(λ).



Continuous Random Variables

Continuous random variables

- Can assume any value in an interval of real numbers.
- Continuous random variables are uncountable .
- Often "measuring".

Example: height, weight, etc.

Continuous Random Variables: Example

Spin a dial, let X be the angle between the dial and the horizontal (measured in degrees).

- What's the probability for $X \in [0, 180]$?
- What's the probability for $X \in [44.9, 45.1]$?
- What's the probability for $X = 45$?

Continuous Random Variables

X is a continuous random variable.

- We do not talk about the probability of X taking any *individual value*:

$$\mathbb{P}(X = x) = 0, \text{ for any value of } x.$$

- We talk about the probability of X *within a range of values*:

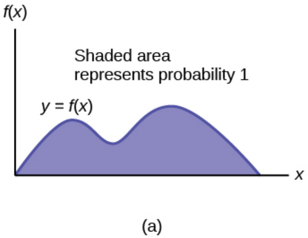
$$\mathbb{P}(X \in [a, b]) \text{ is well defined, for any values of } a, b.$$

The Probability Density Function

The distribution of a continuous random variable is described by a smooth function, $f(x)$, called probability density function (pdf).

- Let X be a continuous random variable with pdf $f(x)$.
- Then the probability that X is between a and b is the *area under the curve* $f(x)$ between $x = a$ and $x = b$.

Figure: The pdf and the area under the curve.



The Probability Density Function

- Using calculus, the probability that X is between a and b is computed as

$$\mathbb{P}(a < X < b) = \int_a^b f(x)dx.$$

- Properties of the probability density function $f(x)$:

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1.$$

* Note: In this class, you will not need to evaluate any integrals.

Expected Value and Variance

Let X be a continuous random variable with pdf $f(x)$.

- The expected value (or mean) of X is

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

- The variance of X is

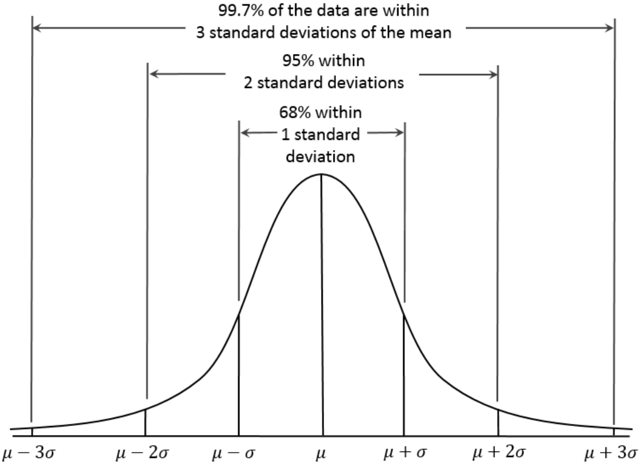
$$\mathrm{var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx.$$

---

* Note: In this class, we will not compute these directly via integrals.

The Normal Distribution

Figure: The pdf of a normal distribution with mean μ and variance σ^2.



The Normal Distribution

The normal distribution is the most important distribution for a continuous random variable. Why is it so important?

- It's mathematically "convenient".
- Many things in the world are approximately normally distributed. If not, can be transformed into normal, e.g. take log.
- (Most important) Sample means tend to have normal distributions, even if the distribution of the population that we're sampling from does not (more on this later).

The Normal Distribution

The pdf for the normal distribution with mean μ and variance σ^2 is:

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

- It is a bell-shaped, continuous function.
- It has two parameters: μ (mean) and σ (standard deviation).
- It is symmetrical about μ and its maximum is at μ.

If X is a normal random variable with mean μ and variance σ^2, we usually write $X \sim \mathcal{N}(\mu, \sigma^2)$.

---

* Note: In this class, we will not work directly with this pdf.

The *Standard* Normal Distribution

Standard normal distribution:

- A normal distribution with $\mu = 0$ and $\sigma = 1$. The simplest normal distribution.
- Denote a standard normal random variable by Z.
- No analytical solution for $\mathbb{P}(a < Z < b)$ for arbitrary a, b.
- People use a precomputed table to find the areas under the curve for Z.

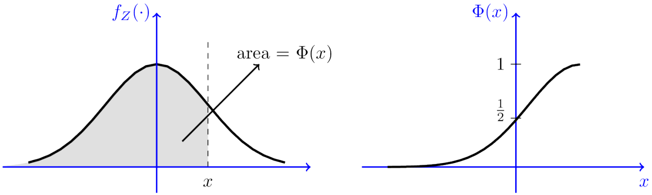
The Cumulative Distribution Function

The cumulative distribution function (CDF) of random variable X is defined as

$$F(x) = \mathbb{P}(X \leq x).$$

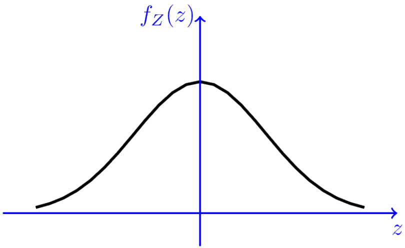
- The CDF of the standard normal distribution is usually denoted by $\Phi(x)$.

Figure: The Φ function (CDF of standard normal Z).



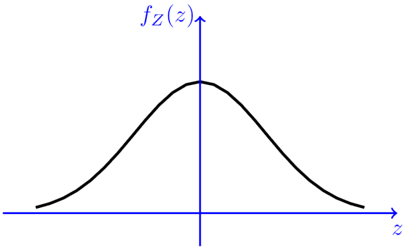
Z-table: the Normal CDF Table

Suppose Z is a standard normal random variable. What is $\mathbb{P}(Z \leq 1.2)$?



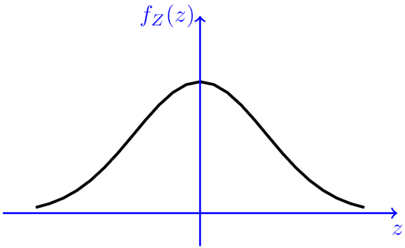
Z-table: the Normal CDF Table

Suppose Z is a standard normal random variable. What is $\mathbb{P}(Z \leq -0.4)$?



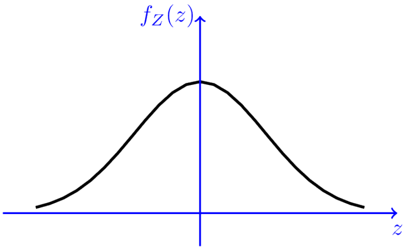
Z-table: the Normal CDF Table

Suppose Z is a standard normal random variable. What is $\mathbb{P}(-0.4 < Z \leq 1.2)$?



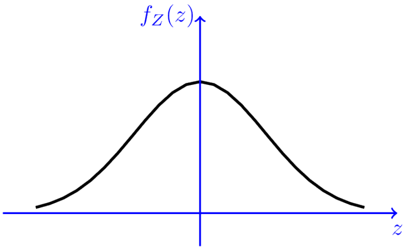
Z-table: the Inverse Normal CDF Table

Suppose Z is a standard normal random variable. Find z such that $\mathbb{P}(Z \leq z) = 0.60$.



Z-table: the Inverse Normal CDF Table

Suppose Z is a standard normal random variable. Find z such that $\mathbb{P}(-z \leq Z \leq z) = 0.95$.



Z-tables

Two tables contain the same information.

- Normal CDF Table: the cutoff *z values* are rounded. It is used to look up for probabilities.
- Inverse Normal CDF Table: the *probabilities* are rounded. It is used to look up for cutoff z values.

Convert Normal to Standard Normal

Shifting and rescaling property.

- If X is a *normal* with mean μ and standard deviation σ, and a, b are two constants. Then, $aX + b$ is a *normal* with mean $a\mu + b$ and standard deviation $|a|\sigma$.

$$X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2).$$

- In particular, $Z = \frac{X-\mu}{\sigma}$ is a standard normal:

$$\mathbb{E}\left(\frac{X-\mu}{\sigma}\right) = \frac{\mu-\mu}{\sigma} = 0,$$
$$\mathrm{sd}\left(\frac{X-\mu}{\sigma}\right) = \frac{\sigma}{\sigma} = 1.$$

Summary

Continuous random variables

- Probability density function (pdf)
- Area under the curve

Normal distribution

- Normal distribution's pdf
- Standard normal distribution, z-tables
- Convert normal to standard normal