Mathematical statistics

August 30th, 2018

Lecture 2: Working with normal distributions

Overview

- Continuous random variables
- Distribution functions
- Working with the standard normal distribution $\mathcal{N}(0,1)$
- Working with the normal distributions $\mathcal{N}(\mu,\sigma^2)$
- Linear combination of normal random variables

Reading: Sections 4.1, 4.2, 4.3

Continuous random variables

Definition

Let X be a random variable. Suppose that there exists a nonnegative real-valued function $f : \mathbb{R} \to [0, \infty)$ such that for any subset of real numbers A, we have

$$P(X \in A) = \int_A f(x)dx$$

Then X is called **absolutely continuous** or, for simplicity, **continuous**. The function f is called the **probability density function**, or simply the **density function** of X.

Whenever we say that X is continuous, we mean that it is absolutely continuous and hence satisfies the equation above.

Properties

Let X be a continuous r.v. with density function f, then

- $f(x) \geq 0$ for all $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x)\ dx = 1$
- For any fixed constant a, b,

$$P(a \leq X \leq b) = \int_a^b f(x)\ dx$$

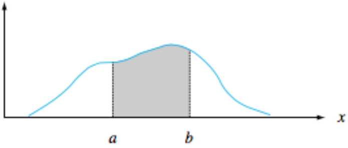


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Expectation

Definition *If X is a continuous random variable with probability density function f, the* ***expected value*** *of X is defined by*

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

The expected value of X is also called the **mean**, or **mathematical expectation**, or simply the **expectation** of X, and as in the discrete case, sometimes it is denoted by EX, $E[X]$, μ, or μ_X.

Theorem 6.3 *Let X be a continuous random variable with probability density function $f(x)$; then for any function $h : \mathbf{R} \to \mathbf{R}$,*

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

Example

Problem

Let X be a continuous r.v. with density function

$$f(x) = \begin{cases} \frac{1}{2}x & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

where c is some unknown constant.

- *Compute $P(X \in [0.25, 0.75])$*
- *Compute $E[X]$ and $Var(X)$.*

Definition

If X is a random variable, then the function F defined on $(-\infty, \infty)$ by

$$F(t) = P(X \leq t)$$

is called the distribution function of X.

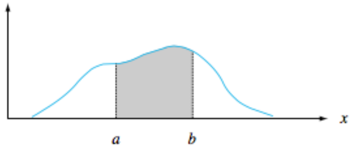


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Distribution function

For continuous random variable:

$$F(t) = P(X \leq t) = \int_{(-\infty,t]} f(x)dx$$
$$= \int_{-\infty}^{t} f(x)dx$$

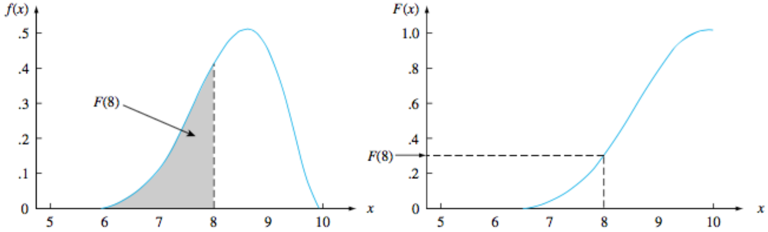


Figure 4.5 A pdf and associated cdf

For continuous random variable:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx = F(b) - F(a)$$

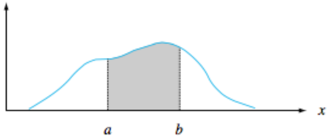


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Moreover:

$$f(x) = F'(x)$$

Example

Problem

The distribution function for the duration of a certain soap opera (in tens of hours) is

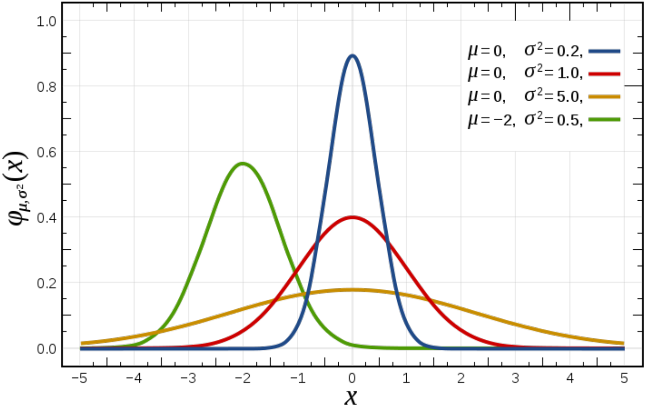
$$F(y) = \begin{cases} 1 - \frac{16}{y^2} & \textit{if } y \geq 4 \\ 0 & \textit{elsewhere} \end{cases}$$

Find $P[4 \leq y \leq 8]$.

Normal random variables

Reading: 4.3

$\mathcal{N}(\mu, \sigma^2)$



$$E(X) = \mu, Var(X) = \sigma^2$$

$\mathcal{N}(\mu, \sigma^2)$

- $E(X) = \mu, \ Var(X) = \sigma^2$
- Density function

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

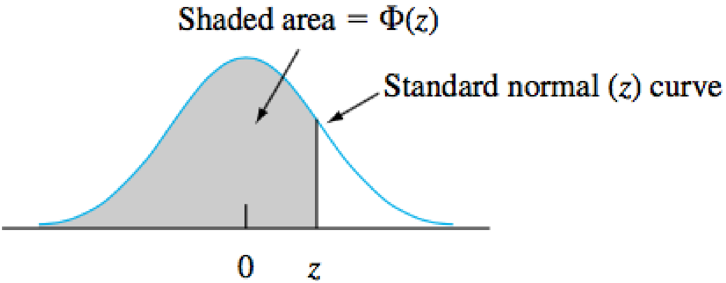
Standard normal distribution $\mathcal{N}(0,1)$

- If Z is a normal random variable with parameters $\mu = 0$ and $\sigma = 1$, then the pdf of Z is

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

  and Z is called the *standard normal distribution*
- $E(Z) = 0, \; Var(Z) = 1$

$\Phi(z)$



$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} f(y)\, dy$$

Table A.3 Standard Normal Curve Areas (*cont.*) $\Phi(z) = P(Z \leq z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

Exercise 1

Problem

Let Z be a standard normal random variable.
Compute

- $P[Z \leq 0.75]$
- $P[Z \geq 0.82]$
- $P[1 \leq Z \leq 1.96]$
- $P[Z \leq -0.82]$

Note: The density function of Z is symmetric around 0.

Exercise 2

Problem

Let Z be a standard normal random variable. Find a, b such that

$$P[Z \leq a] = 0.95$$

and

$$P[-b \leq Z \leq b] = 0.95$$

$\mathcal{N}(\mu, \sigma^2)$

- $E(X) = \mu,\ Var(X) = \sigma^2$
- Density function

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

Shifting and scaling normal random variables

Problem

Let X be a normal random variable with mean μ and standard deviation σ.
Then

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

follows the standard normal distribution.

Shifting and scaling normal random variables

If X has a normal distribution with mean μ and standard deviation σ, then

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

has a standard normal distribution. Thus

$$\begin{aligned} P(a \le X \le b) &= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

$$P(X \le a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \quad P(X \ge b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Exercise 3

Problem

Let X be a $\mathcal{N}(3,9)$ random variable. Compute $P[X \leq 5.25]$.

Linear combination of normal random variables

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent normal random variables (with possibly different means and/or variances). Then

$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

also follows the normal distribution.

Linear combination of normal random variables

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent normal random variables (with possibly different means and/or variances). Then

$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

also follows the normal distribution.

What are the mean and the standard deviation of T?

- $E(T) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$
- $\sigma_T^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + \ldots + a_n^2\sigma_{X_n}^2$

Example 1

Problem

Assume that

$$X_1 \sim \mathcal{N}(10, 9) \qquad \textit{and} \qquad X_2 \sim \mathcal{N}(30, 16)$$

are independent.

What is the distribution of $X_1 - X_2$*?*

Example 2

Problem

A concert has three pieces of music to be played before intermission. The time taken to play each piece has a normal distribution.
Assume that the three times are independent of each other. The mean times are 15, 30, and 20 min, respectively, and the standard deviations are 1, 2, and 1.5 min, respectively.

What is the distribution of the length of the concert?