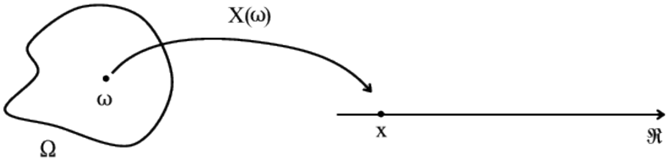
MATH 205: Statistical methods

Lecture 16: Distribution function

Random variables



Notations:

- random variables are denoted by uppercase letters (e.g., X);
- the observed values of the random variables are denoted by lowercase letters (e.g., x)

Discrete random variable

A random variables X is discrete if the set of all possible values of X

- is finite
- is countably infinite

A random variables is characterized by its probability mass function, usually represented as a table

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02

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

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,

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

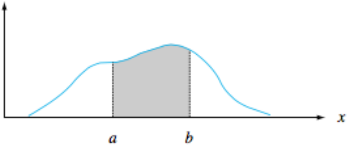


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Distribution function

Definition

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

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

is called the distribution function of X.

Example

Problem

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

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

Compute

- $P[X \leq -1]$
- $P[X \leq 0.5]$
- $P[X \leq 1.5]$

Example

Problem

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

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

- *Let t be a number such that $0 < t < 1$, compute $P[X \leq t]$ (the result should be a function of t)*
- *Find t such that $P[X \leq t] = 1/2$*

Distribution function

For continuous random variable:

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

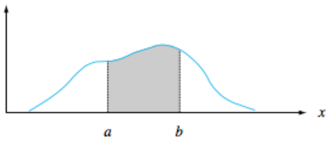


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

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 $P[4 \leq Y \leq 8]$.

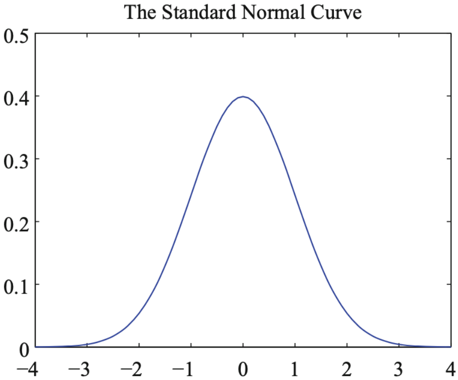
The standard normal distribution

Definition

The probability density function

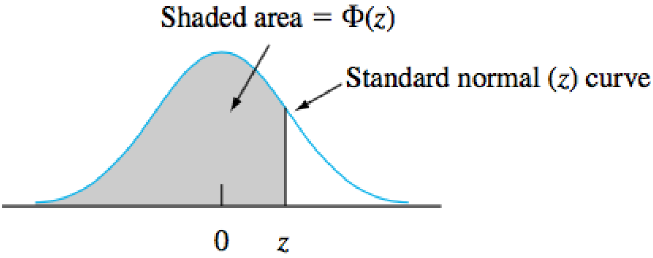
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

is known as the standard normal distribution.



The standard normal distribution has mean 0 and variance 1.

$\Phi(z)$: Distribution function of standard normal



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

$\Phi(z)$

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

Problem

Let Z be a standard normal random variable. Compute

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