MATH 205: Statistical methods

October 11th, 2021

Lecture 11: Uniform and normal distributions

Announcements

- Homework 3 was uploaded to the course webpage, due next Wednesday
- There will be a quiz in class next Monday. The quiz covers the materials of Chapter 4.
- Countdown to midterm exam (written part): 16 days. The exam covers everything up to Chapter 6.
- Countdown to midterm exam (simulations): 14 days (050L) or 16 days (051L). The exam covers everything up Central Limit Theorem.

Revised schedule

Date	Theme/Topic	Labs	Assignments
Sep 1	Syllabus		
Sep 8	Chapter 1: Describing dataset	Section 2: Handling data	
Sep 13 - 15	Chapter 2: Looking at Relationships	Section 3: Univariate data	
Sep 20-22	Chapter 3: Basic Ideas in Probability	Section 4: Bivariate Data	
Sep 27-29	Chapters 3-4	Random data	Homework 1 (due 09/22)
Oct 4-6	Chapter 4: Random variables and expectations	Random continuous data	Homework 2 (due 10/06)
Oct 11-13	Chapter 5: Useful distributions	Section 7: The central limit theorem	
Oct 18-20	Chapter 6: Samples and populations	Section 9: Confidence interval estimation	Homework 3 (due 10/20)
Oct 25-27	Review and midterm exam		Midterm: Oct 27 (lecture), Oct 25-27 (labs)
Nov 1-3	Chapter 7: The significance of evidence	Section 10: Hypothesis testing	
Nov 8-10	Goodness of Fit	Section 12: Goodness of Fit	Homework 4 (due 11/10)
Nov 15-17	Linear Regression	Section 13: Linear regression	
Nov 22-24	Thanksgiving break		
Nov 29 - Dec 1	One-Way Analysis of Variance	Section 15: Analysis of variance	Homework 5 (due 12/01)
Dec 6-8	Selected topics + Review		
Exam week			

Review: Random variables and expectations

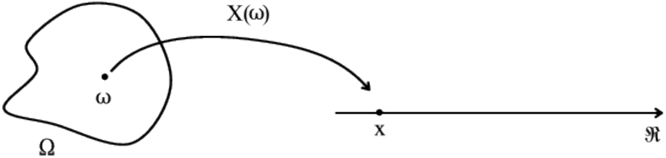
Random variables and expectations

4.1 Random variables

4.2 Expectations

- Expected values
- Mean, variance and covariance

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)

Random variables

- Discrete

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

- Continuous

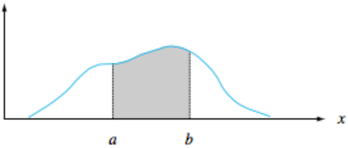


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Expected value: discrete variables

Definition

Given a discrete random variable X which takes values in the set $\mathcal{D}$ and which has probability distribution P:

- The expected value of X is defined as

$$\mathbb{E}[X] = \sum_{x \in \mathcal{D}} x P(X = x)$$

- $f(X)$ is a also discrete random variable, which we write F and

$$\mathbb{E}[f(X)] = \sum_{u \in \mathcal{D}_f} u P(F = u) = \sum_{x \in \mathcal{D}} f(x) P(X = x)$$

  which is sometimes referred to as "the expectation of f".

Expected value: continuous variables

Definition

Given a discrete random variable X which takes values in the set $\mathcal{D}$ and which has probability density function $p(x)$:

- The expected value of X is defined as

$$\mathbb{E}[X] = \int_{\mathcal{D}} xp(x)\ dx$$

- $f(X)$ is also a continuous random variable, which we write F and

$$\mathbb{E}[f(X)] = \int_{\mathcal{D}} f(x)p(x)\ dx$$

  which is sometimes referred to as "the expectation of f"

Mean and variance

Definition

- The mean or expected value of a random variable X is

$$\mathbb{E}[X]$$

- The variance of a random variable X is

$$var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

- The standard deviation of a random variable X is defined as

$$std(X) = \sqrt{var(X)}$$

Expectations are linear

Useful Facts 4.2 (Expectations Are Linear)

Write f, g for functions of random variables.

- $\mathbb{E}[0] = 0$
- for any constant k, $\mathbb{E}[kf] = k\mathbb{E}[f]$
- $\mathbb{E}[f + g] = \mathbb{E}[f] + \mathbb{E}[g]$.

Properties of variance

Lemma

We have

- *For any constant* k, $var[k] = 0$;
- $var[X] \geq 0$;
- $var[kX] = k^2 var[X]$

The variance of a random variable X *can be computed by*

$$var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Covariance

Definition

The covariance of of two random variables X and Y is

$$cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

The covariance of of two random variables X and Y can be computed as

$$cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Independent variables have zero covariance

Proposition

If X and Y are independent, then

- $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
- $cov(X, Y) = 0$
- $var(X + Y) = var(X) + var(Y)$

Example

Example

Assume that the joint probability of X (receive values 1, 2) and Y (receives values 1, 2, 3) is represented by the following table

X \ Y	1	2	3
1	0.14	0.42	0.06
2	0.06	0.28	0.04

Compute Cov(X, Y).

Chapter 5: Useful distributions

The Discrete Uniform Distribution

Definition

A random variable has the discrete uniform distribution if it takes each of k values with the same probability $1/k$, and all other values with probability zero.

Problem

Consider a random variable X that follows discrete uniform distribution on the set $\{1, 2, 3, 4\}$.
Compute $E(X)$ and $Var(X)$.

The continuous uniform distribution

Definition

Write l for the lower bound and u for the upper bound. The probability density function for the uniform distribution on the interval l, u is

$$f(x) = \begin{cases} \frac{1}{u-l}, & x \in [l, u] \\ 0 & otherwise \end{cases}$$

Problem

Consider a random variable X that follows continuous uniform distribution on $[l, u]$.
Compute $E(X)$ and $Var(X)$ (in terms of l, u).

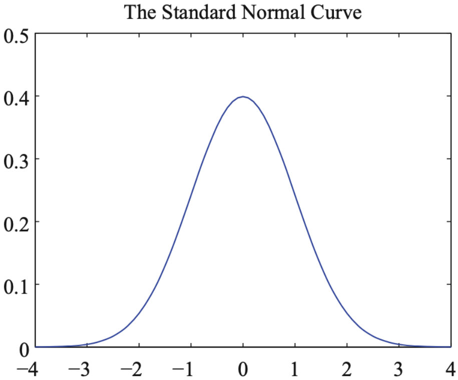
The standard normal distribution

Definition

The probability density function

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

is known as the standard normal distribution.



The standard normal distribution has mean 0 and variance 1.

Normal distributions

Write μ for the mean of a random variable X and σ for its standard deviation; if

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

has a standard normal distribution, then X is a normal random variable.

Definition

The probability density function

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

is a normal distribution.

Normal distributions

Definition

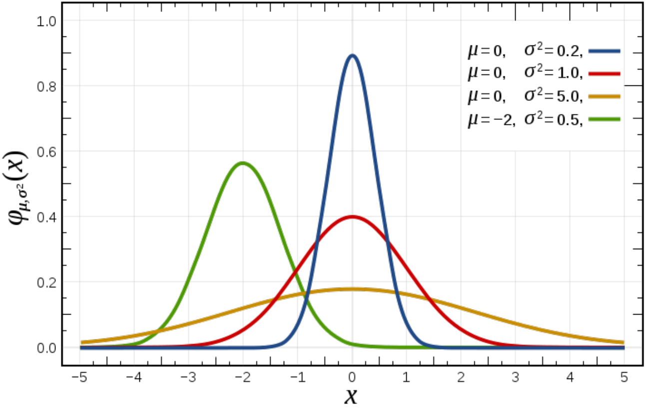
If X a normal random variable with probability density function

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

then

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

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



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

Normal distributions

Questions: If the normal distributions are so complicated, how can we compute probability associated with normal random variables?

Review: Distribution function

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.

Distribution function

For continuous random variable:

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

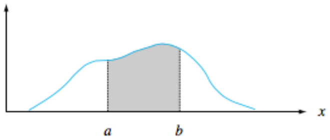
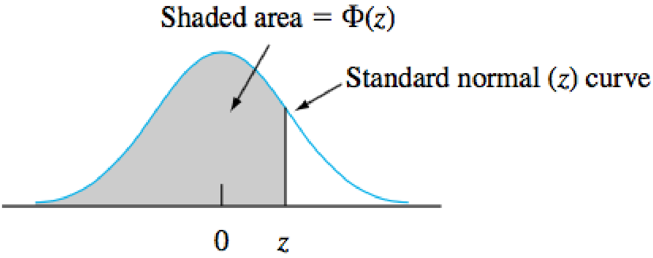


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

$\Phi(z)$: 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 1

Problem

Let Z be a standard normal random variable.
Compute

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

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