MATH 205: Statistical methods

Lecture 23: Distribution of the sample mean

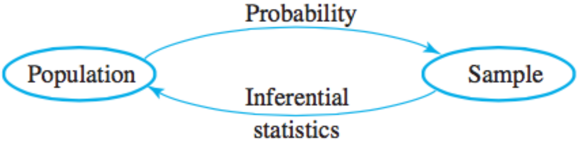
Reminder: Midterm-Written (Friday)

- Closed book.
- Can use calculator.
- Can bring a A4-sized hand-written one-sided note to the exam.

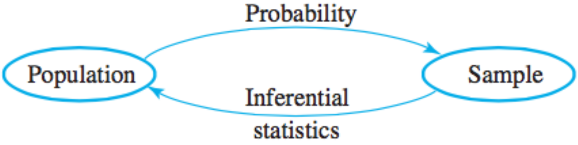
Chapter 6: Samples and Populations

Distributions are like populations

- we can think about population as a probability distribution P
- the samples are random variables X generated from P
- from the observed values of the samples, we want to infer properties about P



Random sample



Definition

The random variables $X_1, X_2, ..., X_n$ are said to form a (simple) random sample of size n if

1. the X_i's are independent random variables
2. every X_i has the same probability distribution

The sample mean is an estimate of the population mean

Definition

Let $X_1, X_2, ..., X_n$ be a random sample from a distribution. The sample mean is defined as

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Questions:

- What can we say about the distribution of $\bar{X}$?
- When can we use $\bar{X}$ to estimate the population mean with confidence?

Reminder: notations

- Let $X_1, X_2, \ldots, X_n$ be a random sample of size n
- The sample mean of $X_1, X_2, \ldots, X_n$, defined by

$$\bar{X} = \frac{X_1 + X_2 + \ldots X_n}{n},$$

  is a random variables
- When the values of $x_1, x_2, \ldots, x_n$ are collected,

$$\bar{x} = \frac{x_1 + x_2 + \ldots x_n}{n},$$

  is a realization of the $\bar{X}$, and is a number

Mean and variance of the sample mean

Theorem

Given independent random samples $X_1, X_2, ..., X_n$ *from a distribution with mean* μ *and standard deviation* σ*, the mean is modeled by a random variable* $\bar{X}$*,*

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Then

$$E[\bar{X}] = \mu$$

and

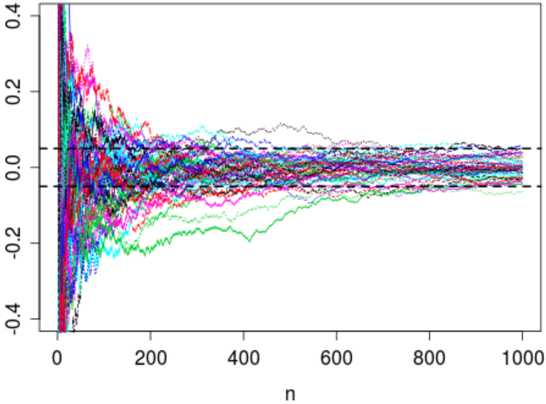
$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Law of large numbers

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then

$$\bar{X} \to \mu$$

as n approaches infinity



The Central Limit Theorem

Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then, in the limit when $n \to \infty$, the $\bar{X}$ follows normal distribution.

Recall that

$$E[\bar{X}] = \mu, \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}},$$

this means we have

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

follows the standard normal distribution.

Rule of Thumb:

If $n > 30$, the Central Limit Theorem can be used for computation.

Example

Problem

Let $X_1, X_2, \ldots, X_{64}$ *be a random sample from a distribution with population mean* $\mu = 1$ *and standard deviation* $\sigma = 2$*.*
Let $\bar{X}$ *be the sample mean*

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_{64}}{64}$$

Compute $P[\bar{X} \leq 1.49]$

$\Phi(z)$

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
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997

Example

Problem

When a batch of a certain chemical product is prepared, the amount of a particular impurity in the batch is a random variable with mean value 4.0 g and standard deviation 1.5 g.

If 50 batches are independently prepared, what is the (approximate) probability that the sample average amount of impurity is between 3.5 and 3.8 g?

Midterm review

Chapter 1& 2: Describing datasets

- Summarizing univariate data
  - mean
  - median
  - standard deviation and variance
  - interquartile range
- Correlation
  - Standard coordinates
  - Using correlation to predict

Chapter 3: Basic ideas in probability

Chapter 4: Random variables and expectations

4.1 Random variables and probability distribution

- Discrete
- Continuous
- Joint and marginal distributions
- Independent variables

4.2 Expectations

- Mean
- Variance
- Covariance

Chapter 5 & Chapter 6

- Working with normal random variables
- Linear combinations of random variables
- ~~Distribution of the sample mean~~