MATH 205: Statistical methods

October 13th, 2021

Lecture 12: Useful distributions

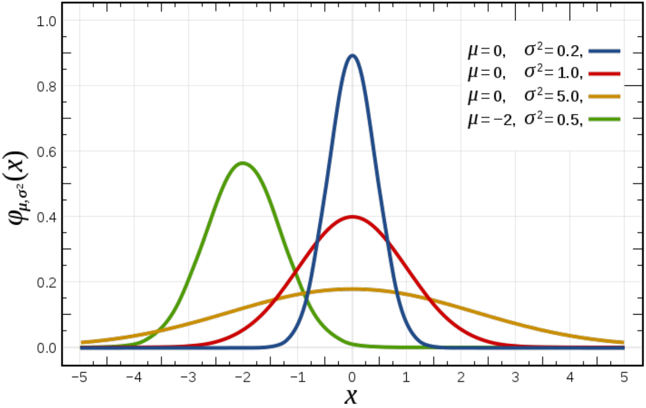
Announcements

- 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.

Last lecture

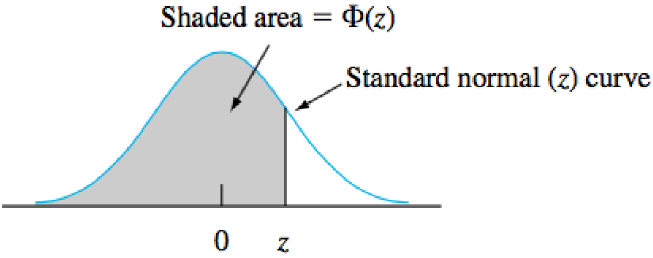
- The discrete uniform distribution
- The continuous uniform distribution
- Normal distributions

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



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

$\Phi(z)$: 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

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

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.

Exercise 2

Problem

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

This lecture: Other useful distributions

- Bernoulli random variables
- The Binomial Probability Distribution
- The Geometric Distribution
- The Poisson Distribution
- The Beta Distribution
- The Gamma Distribution
- The Exponential Distribution

Bernoulli random variables

Definition

A Bernoulli random variable takes the value 1 with probability p and 0 with probability $1 - p$.
An experiment associated with a Bernoulli random variable is called a Bernoulli trial. p is also called the probability of success.

Problem

Consider a random variable X that follows Bernoulli distribution. Compute $E(X)$ and $Var(X)$ (in terms of p).

The Binomial Probability Distribution

Definition

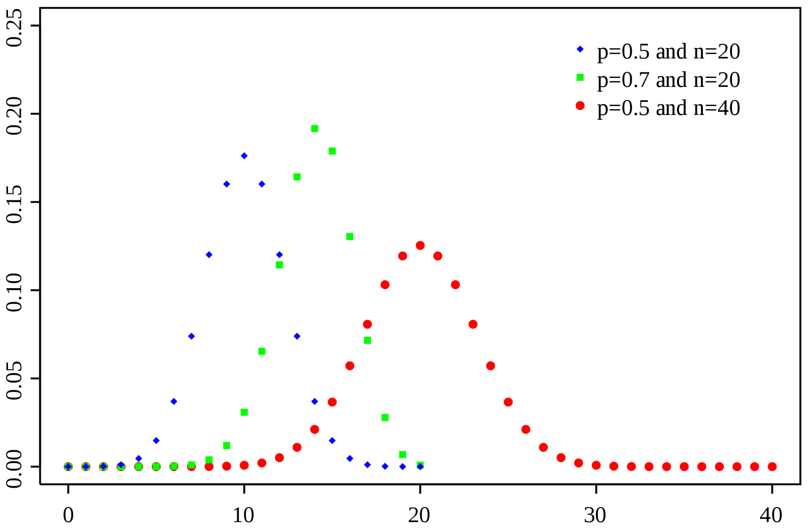
The binomial distribution with parameters N and p is the discrete probability distribution of the number of successes in a sequence of N Bernoulli trials.

$$P(X = n) = \binom{N}{n} p^n (1-p)^{N-n}$$

Recall that:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

The Binomial probability distribution



The Binomial probability distribution

Alternative definition: If $\{X_1, X_2, \dots, X_N\}$ is a sequence of independent Bernoulli random variables with probability p. Then

$$Y = X_1 + X_2 + \dots X_N$$

follows binomial probability distribution $B(N, p)$.

Problem

Consider a random variable X that follows Binomial probability distribution.
Compute $E(X)$ and $Var(X)$ (in terms of p and N).

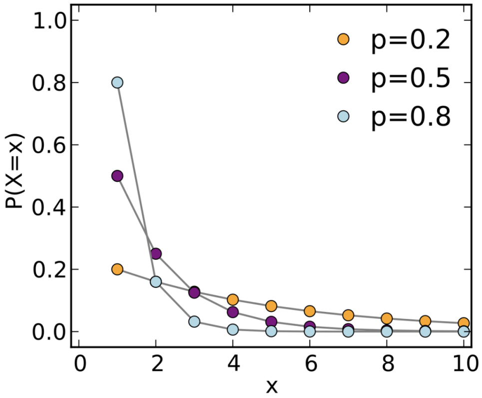
The geometric distribution

Definition

The geometric distribution is the probability distribution of the number X of Bernoulli trials needed to get one success, supported on the set $\{1, 2, 3, \ldots\}$

$$P(X = n) = p(1 - p)^{n-1}$$

The geometric distribution



A geometric distribution with parameter p has mean $1/p$ and variance $(1-p)/p^2$.

Poisson Distribution

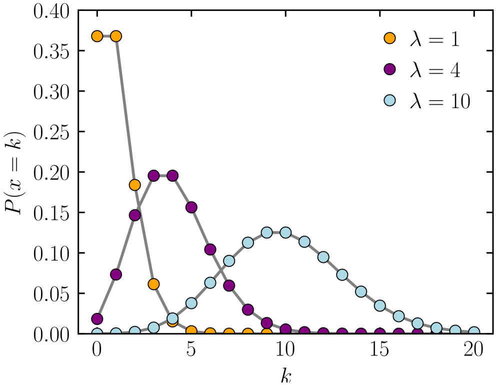
Definition

A non-negative, integer valued random variable X has a Poisson distribution when its probability distribution takes the form

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where $\lambda > 0$ is a parameter often known as the intensity of the distribution.

Poisson Distribution



A Poisson distribution with intensity λ has mean λ and variance λ.

Poisson Distribution

Usually used to model counts that occur in an interval of time or space that

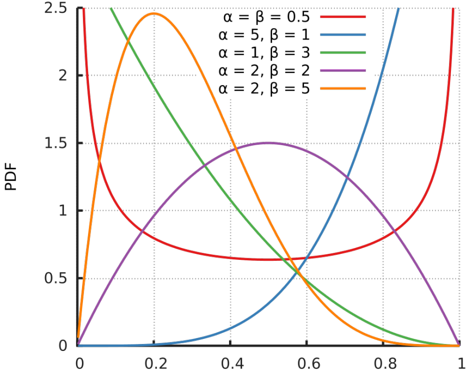
- occur with some fixed average rate
- observation occurs on disjoint interval are independent

Examples:

- the marketing phone calls you receive during the day time
- number of Prussian soldiers killed by horse-kicks each year
- the number of raisins in a loaf/slice of raisin bread

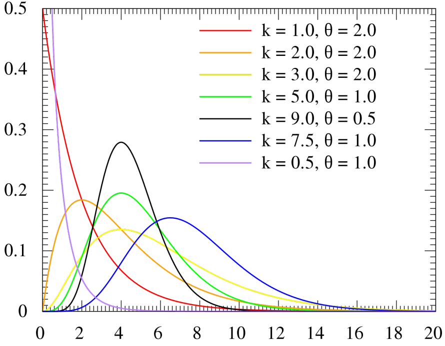
Beta distributions

The Beta distribution is a family of continuous probability distributions defined on the interval $[0, 1]$ parameterized by two positive shape parameters, denoted by α and β, that control the shape of the distribution.



Gamma distributions

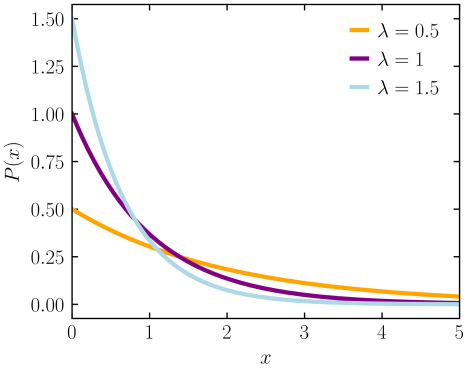
The Beta distribution is a family of continuous probability distribution for a non-negative continuous random variable, parameterized by two positive shape parameters, denoted by α and β, that control the shape of the distribution.



Exponential distributions

A special case of Gamma is the exponential distribution ($\alpha = 1$)

$$f(x) = \beta e^{-\beta x}, \qquad x > 0$$



Linear combinations of random variables

Linear combination of random variables

Theorem

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

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

then the mean and the standard deviation of T can be computed by

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

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

$$T = a_1 X_1 + a_2 X_2 + \ldots a_n X_n$$

also follows the normal distribution with

- $E(T) = a_1 E(X_1) + a_2 E(X_2) + \ldots + a_n E(X_n)$
- $Var(T) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + \ldots + a_n^2 Var(X_n)$

Example

Problem

A gas station sells three grades of gasoline: regular unleaded, extra unleaded, and super unleaded. These are priced at 2.20, 2.35, and 2.50 per gallon, respectively.
Let X_1, X_2, and X_3 denote the amounts of these grades purchased (gallons) on a particular day. Suppose the X_i's are independent with $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$, $\sigma_3 = 50$. Compute the expected value and the standard deviation of the revenue from sales

$$Y = 2.2X_1 + 2.35X_2 + 2.5X_3.$$

Mean and variance of the sample mean

Problem

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 $\bar{X}$,

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

Compute $E(\bar{X})$ and $Var(\bar{X})$(in terms of μ and σ.)

Example

Problem

Let $X_1, X_2, \ldots, X_n$ be independent random samples from $\mathcal{N}(\mu, \sigma^2)$ (that is, normal distribution with mean μ and standard deviation σ).

Let $\bar{X}$ be the sample mean

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

What is the distribution of $\bar{X}$?

Example

Problem

When a batch of a certain chemical product is prepared, the amount of a particular impurity in the batch is a normal 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 X is between 3.5 and 3.8 g?

Hint:

- First, compute $E(\bar{X})$ and $\sigma_{\bar{X}}$
- Note that

$$\frac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}}$$

  is standard normal.

Example

Problem

The tip percentage at a restaurant follows normal distribution with mean value of 18% and a standard deviation of 6%.

What is the approximate probability that the sample mean tip percentage for a random sample of 40 bills is between 16% and 19%?