List of definitions and theorems

I.1 Binomial Distribution

THEOREM 1

Denote by X a random variable counts the number of success. Then the probability of k successes ($k = 0, 1, 2, \ldots, n$) in n Bernoulli trials is

$$Pr(X = k) = \binom{n}{k} p^k \cdot (1 - p)^{n-k}.$$

THEOREM 2 The mean can be calculated by

$$\mu = n \cdot p, \tag{1}$$

PROOF:

Since $X = Y_1 + Y_2 + \ldots + Y_n$ is a sum of indicators (i.e. $Y_j = 1$ at success of j^{th} Bernoulli trial and $Y_j = 0$ otherwise) thus

$$E(X) = E(Y_1) + E(Y_2) + \ldots + E(Y_n).$$

Furthermore clearly for each $j = 1, 2, \ldots, n$

$$E(Y_j) = 1 \cdot p + 0 \cdot (1 - p) = p,$$

we obtain

$$E(X) = n \cdot p.$$

I.2 The Hypergeometric Distribution

THEOREM 1

The probability of k successful selections is then

$$Pr(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

$(k = 0, 1, 2, \dots, n,\ n \leq M,\ n \leq N - M)$

MEAN, VARIANCE

Take n samples and let X_i equal 1 if selection i is successful and 0 if it is not. The i^{th} selection has an equal likelihood of being in any trial, so the chance of a success is

$$p = \frac{M}{N},$$

i.e.

$$Pr(X_i = 1) = \frac{M}{N}.$$

Using the fact

$$E(X_1 + X_2 + \dots X_n) = E(X_1) + E(X_2) + \dots E(X_n),$$

we get that the expectation value of this distribution is

THEOREM 2

The expectation value (the mean) of a hypergeometric distribution is

$$\mu = \frac{nM}{N} = np,$$

where p indicates the probability of getting "good" ball.

(Indeed from N balls M are good.)

Without proof we state that the variance can be calculated as

THEOREM 3

The variance of a hypergeometric distribution is

$$V(X) = npq(1 - \frac{n-1}{N-1}),$$

where q as usual $q = 1 - p$.

I.2 The Poisson Distribution

DEFINITION: A distribution of a random variable X is said to be Poisson Distribution if

$$Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad (A)$$

where $\lambda > 0$ and $k = 0, 1, 2 \ldots$.

THEOREM 1

The above mentioned function $f(k) = Pr(X = k)$ is a probability function indeed.

PROOF:

We have to check:

1.) $f(k) \geq 0$ for every $k = 0, 1, 2, \ldots$

2.) $\sum_k f(k) = 1$.

The first is clearly holds (each factor is positive). The second;

$$\sum_k f(k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}.$$

But from the third tools of the calculus we know that $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$, substituting λ into x, we obtain that

$$e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{1\lambda} = 1,$$

as we wanted.

THEOREM 2

The mean (expected number) is $E(X) = \mu = \lambda$.

PROOF:

By the definition

$$E(X) = \mu = \sum_k k f(k) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} =$$

simplifying by k and taking λ

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda \cdot e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} =$$

substitute a new variable n instead of $k - 1$

$$= \lambda \cdot e^{-\lambda} \sum_{n=0}^{\infty} k \cdot \frac{\lambda^n}{n!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} =$$

$$\lambda,$$

which proves the theorem.

THEOREM 3

The variance is $V(X) = \mu = \lambda$.

PROOF:

By the definition we obtain

$$V(X) = E(X^2) - E(X)^2 = \sum_k k^2 f(k) - \lambda^2 = \sum_{k=1}^{\infty} k^2 \cdot \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 =$$

changing k^2 to $k^2 - k + k$,

$$\sum_{k=1}^{\infty} (k^2 - k + k) \cdot \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 =$$

$$\sum_{k=1}^{\infty} (k^2 - k) \cdot \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 =$$

the second term is λ,

$$= \sum_{k=2}^{\infty} k(k-1) \cdot \frac{\lambda^k}{k!} e^{-\lambda} + \lambda - \lambda^2 =$$

and simplifying by $k(k-1)$ we get

$$= \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} e^{-\lambda} + \lambda - \lambda^2 =$$

$$= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda - \lambda^2 =$$

let the new variable $n = k - 2$

$$= \lambda^2 \sum_{n=0}^{\infty} k \cdot \frac{\lambda^n}{n!} + \lambda - \lambda^2 =$$

$$= \lambda^2 \cdot 1 + \lambda - \lambda^2 = \lambda,$$

so $V(X) = \lambda$.

Continuous random variables probability distribution;

II.1 Density function

DEFINITION:

A $Continuous\ Random\ Variable$ is a random variable whose set of possible values (range) is an interval on the real line. This interval may be open, semiopen, or closed, and it may be bounded or unbounded.

DEFINITION:

The function $f(x)$ is said to be a probability density function (or briefly density function) for a continuous random variable if hold the following three assumption:

1. $f(x) \geq 0$ for all $x \in (-\infty, \infty)$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. The probability that the random variable X lies in the interval (a, b) is given by

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

II.1 Distribution Function

If X is a random variable with probability density function $f(x)$ then the associated cumulative distribution function is defined by the following definition

DEFINITION:

$$F(z) := Pr(X \leq z) = \int_{-\infty}^{z} f(x)dx.$$

From this definition we obtain that

$$Pr(A \leq X \leq B) = \int_{-\infty}^{B} f(x)dx - \int_{-\infty}^{z} Af(x)dx =$$

$$\int_A^B f(x)dx.$$

Practically we can use this observation in two ways;

If the cumulative distribution function is given then

$$Pr(A \leq X \leq B) = F(B) - F(A)$$

If the probability density function is given then

$$Pr(A \leq X \leq B) = \int_A^B f(x)dx.$$

Now we list the most important properties of the cumulative distribution function:

THEOREM:

If $f(x)$ is a probability density function then

1.) $F(z) = \int_{-\infty}^{z} f(x)dx$,

2. $F'(x) = f(x)$,

3.) For every $x \in (-\infty, \infty)$, we have $0 \leq F(x) \leq 1$,

4.) $F(x)$ is non-decreasing,

5.) $\lim_{x\to\infty} F(x) = 1$.

II.1 Expected number (Mean), Variance, Standard deviation of Random Variables

The definition of mean, variance and standard deviation

DEFINITION:

Let $f(x)$ be the probability density function for a continuous random variable X.

1. The mean of X is

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

assuming that the improper integral

$$\int_{-\infty}^{\infty} |x|f(x)dx$$

exists, and convergent.

2.The variance of X is

$$V(X) = E((X - \mu)^2)$$

3. The standard deviation of X is

$$\sigma = \sqrt{V(x)}.$$

THEOREM 1:

$$V(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2.$$

PROOF:

By the definition and the well-known identity

$$V(X) = E((X - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx =$$

$$= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x) dx =$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx - \int_{-\infty}^{\infty} 2x \cdot \mu \cdot f(x) dx +$$

$$+ \int_{-\infty}^{\infty} \mu^2 f(x) dx =$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx - 2\mu \int_{-\infty}^{\infty} x f(x) dx +$$

$$+ \mu^2 \int_{-\infty}^{\infty} f(x) dx.$$

Observe that $\int_{-\infty}^{\infty} f(x)dx = 1$ and that the middle integral is μ. Therefore

$$V(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - 2\mu \cdot \mu + \mu^2 =$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2.$$

Uniform and Exponential Distributions

UNIFORM DISTRIBUTION

The uniform distribution has a constant success rate on the interval $a \leq x \leq b$ and zero success rate anywhere else. Therefore we conclude that its probability density function is the following type:

$$f(x) = \begin{cases} c & \text{, if } a \leq x \leq b \\ 0 & \text{, otherwisw} \end{cases}.$$

Now we shall determine the constant c, using the definition of the density function. Since $f(x) \geq 0$, we obtain that $c \geq 0$. Furthermore we should have

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_a^b c dx = [cx]_a^b = c(b-a),$$

which implies that

$$c = \frac{1}{b-a}.$$

Thus we obtain

THEOREM:

The probability density function of a uniformly distributed random variable is

$$f(x) = \begin{cases} \frac{1}{b-a} & , \text{if } a \leq x \leq b \\ 0 & , \text{otherwisw} \end{cases} .$$

Now we can derive the cumulative distribution function as well.

THEOREM:

The cumulative distribution function of a uniformly distributed random variable is

$$F(x) = \begin{cases} 0 & , \text{if } x < a \\ \frac{x-a}{b-a} & , \text{if } a \leq x \leq b \\ 1 & , \text{if } x > b \end{cases} .$$

PROOF:

We split the real line into three parts;

1. If $x < a$.

Here $f(x)$ is 0 and thus the improper integral of it (which is the cumulative distribution function) is also 0.

2. If $a \leq x \leq b$.

Then

$$F(x) = \int_a^x \frac{1}{b-a} dt = [t/(b-a)]_a^x = \frac{x-a}{b-a}.$$

Finally

3. If $x > b$. Clearly in this region $F(x)$ is 1.

EXPONENTIAL DISTRIBUTION

Exponential distribution is probably the most widely known and used distribution. Exponential distribution is frequently used for the analysis of time-dependent data when the rate at which events occur does not vary.

The exponential distribution is characterized by the following probability density function:

DEFINITION:

Let $\lambda > 0$ be a real number the random variable X is said to be exponentially distributed if its probability density function is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , 0 \leq x \\ 0 & , \text{otherwise} \end{cases} .$$

THEOREM:

The above mentioned function $f(x)$ is probability density function.

PROOF:

1. The first condition is trivial; λ positive, $e^{-\lambda x}$ is positive as well, hence $f(x) \geq 0$.

2. We need

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Now

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \lambda e^{-\lambda x}dx =$$

$$= [e^{-\lambda x}]_0^{\infty} = [1 - \lim_{x\to\infty} e^{-\lambda x}] = 1 - 0 = 1.$$

THEOREM:

Let X be an exponentially distributed random variable. Its cumulative distribution function is

$$F(x) = \begin{cases} \text{1-e}^{-\lambda x} & , 0 \leq x \\ 0 & , \text{otherwise} \end{cases}.$$

PROOF:

For negative x the density function is 0, hence the cumulative distribution function is also 0.

So let us assume that $x \geq 0$. Then

$$F(x) = \int_0^x \lambda e^{-\lambda t}dt = [e^{-\lambda t}]_0^x = [1 - e^{-\lambda x}].$$

MEAN, VARIANCE, STANDARD DEVIATION OF THE EXPONENTIAL DISTRIBUTION

THEOREM:

If X is exponentially distributed random variable, then

$$\mu = E(X) = \frac{1}{\lambda},$$

and the variance is

$$V(X) = \frac{1}{\lambda^2}.$$

PROOF:

By the definition we have

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} x\lambda e^{-\lambda x}dx.$$

First we shall treat the antiderivative

$$\lambda e^{-\lambda x} dx = \lambda^{-\lambda x} dx.$$

Using the "Integration by parts" we calculate $^{-} dx$ as follows:

Let $u(x) = x$; and $v'(x) = e^{-\lambda x}$. Then $u'(x) = 1$; and $v(x) = \frac{-1}{\lambda} e^{-\lambda x}$. Hence

$$\int x e^{-\lambda x} dx = \frac{-1}{\lambda} x e^{-\lambda x} + \int \frac{1}{a} e^{-\lambda x} dx =$$

$$= \frac{-1}{a} x e^{-\lambda x} - \frac{1}{\lambda^2} e^{-\lambda x}.$$

Thus (using the L'Hospital's rule as well)

$$\mu = E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = \lim_{T\to\infty} \int_0^T x\lambda e^{-\lambda x} dx =$$

$$= \lim_{T\to\infty} [\frac{-1}{\lambda} x e^{-\lambda x} - \frac{1}{\lambda^2} e^{-\lambda x}]_0^T =$$

$$= \lim_{T\to\infty} [-T/e^{\lambda T} - 1/(\lambda e^{\lambda T}) + \frac{1}{\lambda} e^0] =$$

$$= \frac{1}{\lambda}.$$

In the rest we shall calculate the variance.

$$V(X) = \int_{-\infty}^\infty x^2 f(x) dx - \frac{1}{\lambda^2}.$$

Our task is to calculate the antiderivative

$$\int x^2 f(x) dx = \int x^2 e^{-\lambda x} dx.$$

Using the "Integration by parts" again, let $u(x) = x^2$; and $v'(x) = e^{-\lambda x}$. Then $u'(x) = 2x$; and $v(x) = \frac{-1}{\lambda} e^{-\lambda x}$. Thus

$$\int x^2 f(x) dx = \int x^2 e^{-\lambda x} dx = \frac{-1}{\lambda} x^2 e^{-\lambda x} + \int 2x \frac{1}{\lambda} e^{-\lambda x} =$$

$$= \frac{-1}{\lambda} x^2 e^{-\lambda x} + \frac{2}{\lambda} \int x e^{-\lambda x}.$$

But the integral $\int x e^{-\lambda x}$ is known (we calculated at the mean), so substituting the result we obtain

$$= \int x^2 e^{-\lambda x} dx = e^{-\lambda x} (\frac{-x^2}{\lambda} - \frac{2x}{\lambda}) - \frac{2}{\lambda^2}.$$

Thus - by the L'Hospital's rule- the variance will be

$$V(X) = \lim_{T\to\infty} [-x^2 e^{-\lambda x} - \frac{2x}{\lambda} e^{-\lambda x} - \frac{2}{\lambda^2} e^{-\lambda x}]_0^T - \frac{1}{\lambda^2} =$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2},$$

as we wanted.

Law of large numbers; Markov's, Chebyshev's inequalities

:

THEOREM (THE MARKOV'S THEOREM)

Let X be a random variable and assume that its range is a subset of nonnegative real numbers. Assume that $E(X)$ exists. Furthermore let A be some positive constant, then

$$Pr(X \geq A) \leq \frac{E(X)}{A}.$$

PROOF:

We show this theorem both for discrete and continuous random variables as well.

X is discrete random variable:

The proof is almost trivial. We only have to recall the notion of the mean of a random variable. It is

$$E(X) = \sum_i z_i Pr(X = z_i),$$

where $\{z_i\}$ is the sequence of the range of X. The right hand side contains just nonnegative elements, thus we decrease this sum if we restrict it to those z_i which are $\geq A$. So

$$E(X) = \sum_i z_i Pr(X = z_i) \geq \sum_{z_i \geq A} z_i Pr(X = z_i) \geq$$

$$\geq \sum_{z_i \geq A} A Pr(X = z_i) \geq A \cdot \sum_{z_i \geq A} Pr(X = z_i).$$

We can rewrite $\sum_{z_i \geq A} Pr(X = z_i)$ as $Pr(X \geq A)$, so we have

$$E(X) \geq A \cdot Pr(X \geq A).$$

Dividing by the positive number A we get the theorem in the discrete case.

X has continuous probability density function:

In this case the mean

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x f(x) dx.$$

Since $xf(x)$ is a positive function we decrease the right hand side if we restrict it for $x \geq A$. We obtain

$$\int_0^\infty xf(x)dx \geq \int_{A \leq x < \infty} xf(x)dx$$

and by the "mean-value theorem" of definite integrals (which implies that if $\int_a^b f(x)dx$ and $\int_a^b g(x)dx$ exist and $f(x) \geq A$ on the given integral, then $\int_a^b f(x)g(x)dx \geq \int_a^b A \cdot g(x)dx = A \cdot \int_a^b g(x)dx$) we get

$$E(X) \geq \int_A^\infty xf(x)dx \geq \int_A^\infty Af(x)dx =$$

$$A \int_A^\infty f(x)dx$$

As we have seen in Chapter II $\int_A^\infty f(x)dx = F(A) = Pr(X \geq A)$. Hence

$$E(X) \geq A \cdot Pr(X \geq A).$$

Dividing with positive A we obtain the theorem.

THE CHEBYSHEV'S INEQUALITY:

We shall state Chebyshev's inequality in four (but equivalent) forms

THEOREM

Let X be a random variable with mean and finite variance $V(X)$. Now, for any real number $A > 0$,

$$Pr(|X - \mu| \geq A) \leq \frac{V(X)}{A^2},$$

where $\mu = E(X)$ is the mean (expected value) of the random variable X.

PROOF:

Chebyshev's inequality is a simple consequence of the Markov's inequality. Indeed let us use Markov's inequality for the random variable $Z = (X - \mu)^2$. Clearly its range is a subset of the positive real numbers ($(X - \mu|)^2$ is always nonnegative) Now from the Markov's inequality we get

$$Pr(Z \geq A^2) \leq \frac{E(Z)}{A^2}.$$

But recall (see Chapter III)

$$E(Z) = E((X - \mu)^2) = V(X),$$

and $Z \geq A^2$ is equivalent to $\sqrt{Z} = |X - \mu| \geq A$, Thus

$$Pr(Z \geq A^2) = Pr(|X - \mu| \geq A) \leq \frac{E(Z)}{A^2} = \frac{V(X)}{A^2}.$$

We proved the theorem.

VI.3 EQUIVALENT FORMS OF CHEBYSHEV'S INEQUALITY

First we give the complementary form of Theorem VI.2

THEOREM

With the conditions of Theorem VI.1 we have

$$Pr(|X - \mu| A) \geq 1 - \frac{V(X)}{A^2}.$$

Now let us introduce a parameter $\lambda > 1$ by

$$\lambda = \frac{A}{V(X)}$$

i.e. let

$$A = \lambda \cdot V(X).$$

Substituting $A = \lambda \cdot V(X)$ Theorem VI.2 and Theorem VI.2' we obtain

THEOREM

With the conditions of Theorem VI.1 we have

$$Pr(|X - \mu| \geq \lambda V(X)) \leq \frac{1}{\lambda^2},$$

or in the complementary form

$$Pr(|X - \mu| \leq \lambda V(X)) \geq 1 - \frac{1}{\lambda^2}.$$

THE LAW OF LARGE NUMBERS

DEFINITION: (SAMPLE MEAN)

Repeat an experiment n times and consider the success of a given event A. Assume that all trials are independent. As we have learnt in Chapter I. this is a sequence of Bernoulli trials. Denote the probability of success by p of each trial (probability of failure is then $q = 1-p$) and denote by $\mathbf{x}_n$ the number of occurrence of event A.

We mean on the relative frequency or sample mean of the event A the ratio

$$\frac{\mathbf{x}_n}{n}.$$

THEOREM

For every $\varepsilon > 0$ we have

$$Pr(|\frac{\mathbf{x}_n}{n} - p| \geq \varepsilon) \leq \frac{pq}{\varepsilon^2 n}.$$

We can formulate this theorem in the complementer for as well

THEOREM

For every $\varepsilon > 0$ we have

$$Pr(|\frac{\mathbf{x}_n}{n} - p| \leq \varepsilon) \geq 1 - \frac{pq}{\varepsilon^2 n}.$$

PROOF:

We derive the weak form of law of large numbers from the Chebyshev's inequality.

As we mentioned ξ_n is binomial distributed random variable with mean $\mu = np$ and variance $V(X) = \sqrt{npq}$. Therefore for every $\lambda > 1$ we have

$$Pr(|\xi_n - \mu| \geq \lambda) = Pr(|\xi_n - np| \geq \lambda\sqrt{npq}) \leq \frac{1}{\lambda^2).}$$

The inequality

$$|\xi_n - np| \geq \lambda\sqrt{npq}$$

is equivalent to

$$|\frac{\xi_n}{n} - p| \geq \lambda\sqrt{\frac{pq}{n}}$$

(just dividing by n). Thus we have

$$Pr(|\frac{\xi_n}{n} - p| \geq \lambda\sqrt{\frac{pq}{n}}) \leq \frac{1}{\lambda^2).}$$

Let us denote

$$\varepsilon = \lambda\sqrt{\frac{pq}{n}},$$

and if we express $\frac{1}{\lambda^2}$ from this equality we get

$$\frac{1}{\lambda^2} = \frac{pq}{\varepsilon^2 n}.$$

Plug $\frac{1}{\lambda^2}$ to the Chebyshev's inequality then

$$Pr(|\frac{\xi_n}{n} - p| \geq \varepsilon) \leq \frac{pq}{\varepsilon^2 n},$$

as we wanted.

CENTRAL LIMIT THEOREM

Finally in the rest of this section we state an important result:

THEOREM 7 (Central Limit Theorem)

Let $X_1, X_2, \dots$ be a sequence of independent random variables with the same distributions, i.e. for every n $E(X_n) = \mu_n = \mu$, $\sigma_n = \sigma = \sqrt{V(X_n)}$. Introduce a new random variable

$$Y_n := \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

We have

$$\lim_{n\to\infty} Pr\big(Y_n < A\big) = \Phi(A).$$

NORMAL DISTRIBUTION

DEFINITION:

A continuous random variable X has a normal distribution and it is said to be a normal random variable if its probability density function is the following:

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

DEFINITION:

The normal random variable Z with mean $\mu = 0$ and standard deviation $= 1$ is said to be standard normal random variable, and the graph is called the standard normal curve.

THEOREM:

If X is a normal distribution random variable, with mean μ and standard deviation σ, then we have

$$F(x) = \Phi(\frac{x - \mu}{\sigma}).$$

THEOREM:

For every x we have

$$\Phi(-x) = 1 - \Phi(x).$$