Lecture 2. Distributions and Random Variables

Igor Rychlik

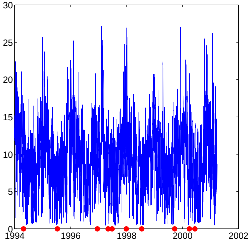
Chalmers

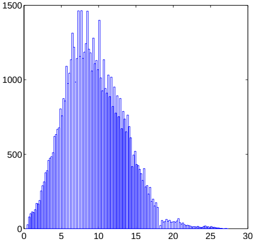
Department of Mathematical Sciences

Probability, Statistics and Risk, MVE300 • Chalmers • March 2011.Click on red text for extra material.

Wind Energy production:

Available Wind Power $p = 0.5\rho_{air} A_r\, v^3$, ρ_{air} air density, A_r area swept by rotor, v - hourly wind speed.





Left: 7 years data. *Right:* Histogram wind "distribution".

Estimate of possible yearly production:

$$p_{yr} = \frac{1}{7} 0.5\rho_{air} A_r \sum_{i=1}^{61354} v_i^3 = 116678, \quad \text{[some units]}.$$

Before age of computers one could estimate p_{yr} using statistics (random variables, law of large numbers): click for details

Random variables:

Often in engineering or the natural sciences, outcomes of random experiments are numbers associated with some physical quantities. Such experiments, called **random variables**, will be denoted by capital letters, e.g., U, X, Y, N, K.

The set $\mathcal{S}$ of possible values of a random variable is a sample space which can be all real numbers, all integer numbers, or subsets thereof.

Example 1 For the experiment flipping a coin, let to the outcomes "Tails" and "Heads" assign the values 0 and 1 and denote by X. One say that X is **Bernoulli distributed**. What does it mean "distributed"?

Probability distribution function:

A statement of the type "$X \leq x$" for any fixed real value x, e.g. $x = -2.1$ or $x = 5.375$, plays an important role in computation of probabilities for statements on random variables and a function

$$F_X(x) = \mathrm{P}(X \leq x), \quad x \in \mathbb{R},$$

is called the **probability distribution**, **cumulative distribution function**, or **cdf** for short.

Example 2 Data, Figures

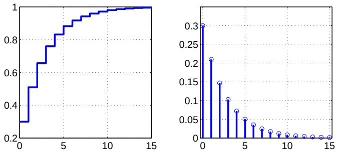
The probability of any statement about the random variable X is computable (at least in theory) when $F_X(x)$ is known.

Example 3

Probability mass function

If K takes a finite or (countable) number of values it is called **discrete** random variables and the distribution function $F_K(x)$ is a "stair" looking function that is constant except the possible jumps. The size of a jump at $x = k$, say, is equal to the probability $\mathrm{P}(K = k)$, denoted by p_k, and called the **probability-mass function**.

Example 4 Pmf



Geometrical distribution with $p_k = 0.70^k \cdot 0.30$, for $k = 0, 1, 2, \ldots$.

Left: Distribution function.

Right: Probability-mass function.

Counting variables

Geometric probability-mass function:

$$P(K = k) = p\,(1 - p)^k, \quad k = 0, 1, 2, \ldots$$

Binomial probability-mass function:

$$P(K = k) = p_k = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

Poisson probability-mass function:

$$P(K = k) = \mathrm{e}^{-m} \frac{m^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Example 5

Ladislaus Bortkiewicz

Ladislaus Bortkiewicz
(1868-1931)

Important book published in 1898:
Das Gesetz der kleinen Zahlen

Law of Small Numbers

If an experiment is carried out by n independent trials and the probability for "success" in each trial is p, then the number of successes K is given by the binomial distribution:

$$K \in \mathrm{Bin}(n, p).$$

If $n \to \infty$ and $p \to 0$ so that $m = n \cdot p$ is constant, we have approximately that

$$K \in \mathrm{Po}(np).$$

(The approximation is satisfactory if $p < 0.1$ and $n > 10$.)

Example 6 Let p be probability that accident occurs during one year, n be number of structures (years) then number of accidents during one year $K \in \mathrm{Po}(np)$, example of accident.

Example 7

CDF - defining properties:

Any function $F(x)$ satisfying the following three properties is a distribution of some random variable:

- The distribution function $F_X(x)$ is non-decreasing function.
- $F_X(-\infty) = 0$ while $F_X(+\infty) = 1$.
- $F_X(x)$ is right continuous.

If $F_X(x)$ is continuous then $\mathrm{P}(X = x) = 0$ for all x and X is called **continuous**. The derivative $f_X(x) = F_X'(x)$ is called **probability density function** (pdf) and

$$F_X(x) = \int_{-\infty}^{x} f_X(z)\, dz.$$

Hence any positive function that integrates to one defines a cdf.

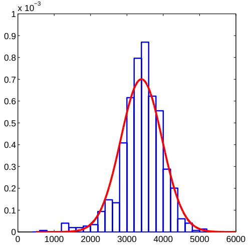
Example 8

Normal pdf- and cdf-function:

The cdf of standard normal cdf is defined through its pdf-function:

$$P(X \leq x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} \, d\xi.$$

The class of normal distributed variables $Y = m + \sigma X$, where m, $\sigma > 0$ are constants is extremely versatile. From a theoretical point of view, it has many advantageous features; in addition, variability of measurements of quantities in science and technology are often well described by normal distributions.



Example 9 Normalized histogram of weights of 750 newborn children in Malmö.

Solid line the normal pdf with $m = 3400$ g, $\sigma = 570$ g.

Is this a good model? Have girls and boys the same weights variability?

Example: Normal cdf - $\Phi(x)$-function:

This table gives function values of $\Phi(x), x \geq 0$. For negative values of x, use that $\Phi(-x) = 1 - \Phi(x)$.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67600	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Classes of distributions - scale and location parameters

For a r.v. X having $F_X(x)$ a random variable $Y = aX + b$ has distribution

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(X \leq (y-b)/a) = F_X((y-b)/a)$$

where a and b are deterministic constants (may be unknown).

If two variables X and Y have distributions satisfying the equation

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right)$$

for some constants a and b, we say that the distributions F_Y and F_X belong to the same **class**; a is called **scale parameter** and b is called **location parameter**.

Standard Distributions

In this course we shall meet many classes of discrete cdf: Binomial, Geometrical, Poisson, ...; and continuous cdf: uniform, normal (Gaussian), log-normal, exponential, χ^2, Weibull, Gumbel, beta ...

Distribution	
Beta distribution, Beta(a,b)	$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1},\ 0<x<1$
Binomial distribution, Bin(n,p)	$p_k=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,\dots,n$
First success distribution	$p_k=p(1-p)^{k-1},\quad k=1,2,3,\dots$
Geometric distribution	$p_k=p(1-p)^k,\quad k=0,1,2,\dots$
Poisson distribution, Po(m)	$p_k=\mathrm{e}^{-m}\frac{m^k}{k!},\quad k=0,1,2,\dots$
Exponential distribution, Exp(a)	$F(x)=1-\mathrm{e}^{-x/a},\quad x\ge 0$
Gamma distribution, Gamma(a,b)	$f(x)=\frac{b^a}{\Gamma(a)}x^{a-1}\mathrm{e}^{-bx},\quad x\ge 0$
Gumbel distribution	$F(x)=\mathrm{e}^{-\mathrm{e}^{-(x-b)/a}},\quad x\in\mathbb{R}$
Normal distribution, N(m,σ^2)	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-(x-m)^2/2\sigma^2},\quad x\in\mathbb{R}$ $F(x)=\Phi((x-m)/\sigma),\quad x\in\mathbb{R}$
Log-normal distribution, $\ln X\in \mathrm{N}(m,\sigma^2)$	$F(x)=\Phi(\frac{\ln x-m}{\sigma}),\quad x>0$
Uniform distribution, U(a,b)	$f(x)=1/(b-a),\quad a\le x\le b$
Weibull distribution	$F(x)=1-\mathrm{e}^{-\left(\frac{x-b}{a}\right)^c},\quad x\ge b$

Quantiles

The α *quantile* x_α, $0 \leq \alpha \leq 1$, is a generalization of the concepts of median and quartiles and is defined as follows:

The **quantile** x_α for a random variable X is defined by the following relations:

$$\mathrm{P}(X \leq x_\alpha) = 1 - \alpha, \quad x_\alpha = F^{-}(1 - \alpha).$$

In some textbooks, quantiles are defined by the relation $\mathrm{P}(X \leq x_\alpha) = \alpha$; then the inverse function $F^{-}(y)$ could be called the "quantile function".

Example 10

Example: Finding λ_α, i.e. quantiles of N(0,1) cdf

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67600	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Independent random variables

The variables X_1 and X_2 with distributions $F_1(x)$ and $F_2(x)$, respectively, are **independent** if for all values x_1 and x_2

$$P(X_1 \leq x_1 \text{ and } X_2 \leq x_2) = F_1(x_1) \cdot F_2(x_2).$$

Similarly, variables $X_1, X_2, \ldots, X_n$ are **independent** if for all $x_1, x_2, \ldots, x_n$

$$P(X_1 \leq x_1,\, X_2 \leq x_2, \ldots,\, X_n \leq x_n) = F_1(x_1) \cdot F_2(x_2) \cdot \ldots \cdot F_n(x_n).$$

If in addition, for all i, $F_i(x) = F(x)$ then $X_1, X_2, \ldots, X_n$ are called **independent, identically distributed** variables (**iid** variables).

Empirical probability distribution

- Suppose experiment was repeated n times rending in a sequence of X values, $x_1, \ldots, x_n$. The fraction $F_n(x)$ of the observations satisfying the condition "$x_i \leq x$"

$$F_n(x) = \frac{\text{number of } x_i \leq x, \ i = 1, \ldots, n}{n}$$

  is called the **empirical cumulative distribution function** (ecdf).

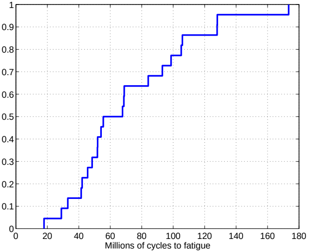
- The **Glivenko–Cantelli Theorem** states that the maximal distance between $F_n(x)$ and $F_X(x)$ tends to zero when n increases without bounds, viz. $\max_x |F_X(x) - F_n(x)| \to 0$ as $n \to \infty$.

- Assuming that $F_X(x) = F_n(x)$, means that the uncertainty in the future (yet unknown) value of X is model by means of drawing a lot from an urn, where lots contain only the observed values x_i. By Glivenko-Cantelli th. this is a good model when n is large.

Example: lifetimes for ball bearings

Data:

17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.48,
51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.88,
84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04,
173.40.



ECDF of ball bearings life time.

Example wind speed data

In this lecture we met following concepts:

- Random variables (rv).
- Probability distribution (cdf), mass function (pmf), density (pdf).
- Law of small numbers.
- Quantiles.
- Empirical cdf.
- You should read how to generate uniformly distributed random numbers.
- How to generate non-uniformly distributed random numbers by just transforming uniform random numbers.

Examples in this lecture "click"