Statistics and Probability I

Bibliography

- Bayesian Data Analysis, Carlin,Stern and Rubin, CHAPMAN & HAA/CRC
- Bayesian Reasoning in Data Analysis, Giulio D'Agnostini, World Scientific.
- ICIC Data Analysis Workshop 2016, Alan Heavens Lectures.
- MACSS 2016 LEcture notes.

- ¿Why do we need a statistics and probability course?

In general

- Infer something from data set
- Test hypothesis.
- Select a model or take decisions.

- ¿Why do we need a statistics and probability course?

In cosmology and astrophysics most of the problems consist of having a set of data from which we want to INFER something.

- Infer some parameter values.
- Test an hypothesis.
- Select a model.

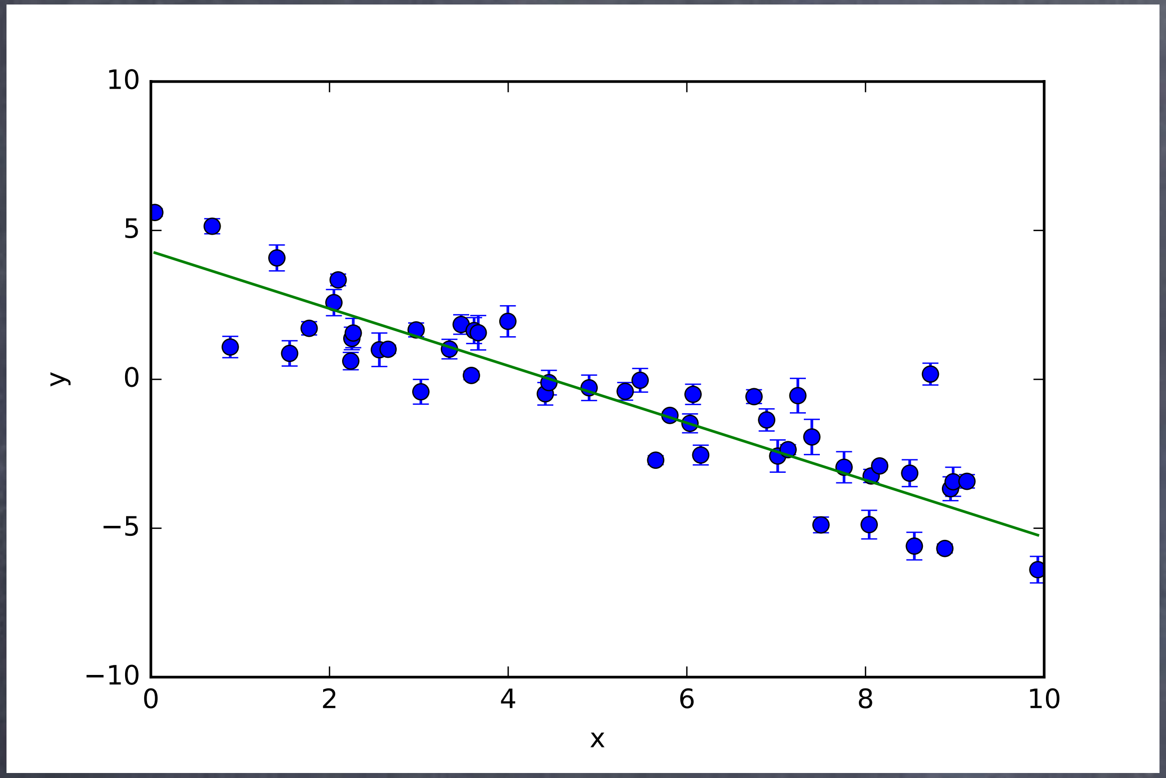
What is the value the parameters involved in the LCDM paradigm?

Is the CMB consistent with a scale free initial power spectrum of fluctations, and with a gaussian distribution?

¿Is General Relativity the correct and final theory, or modified theories works better?

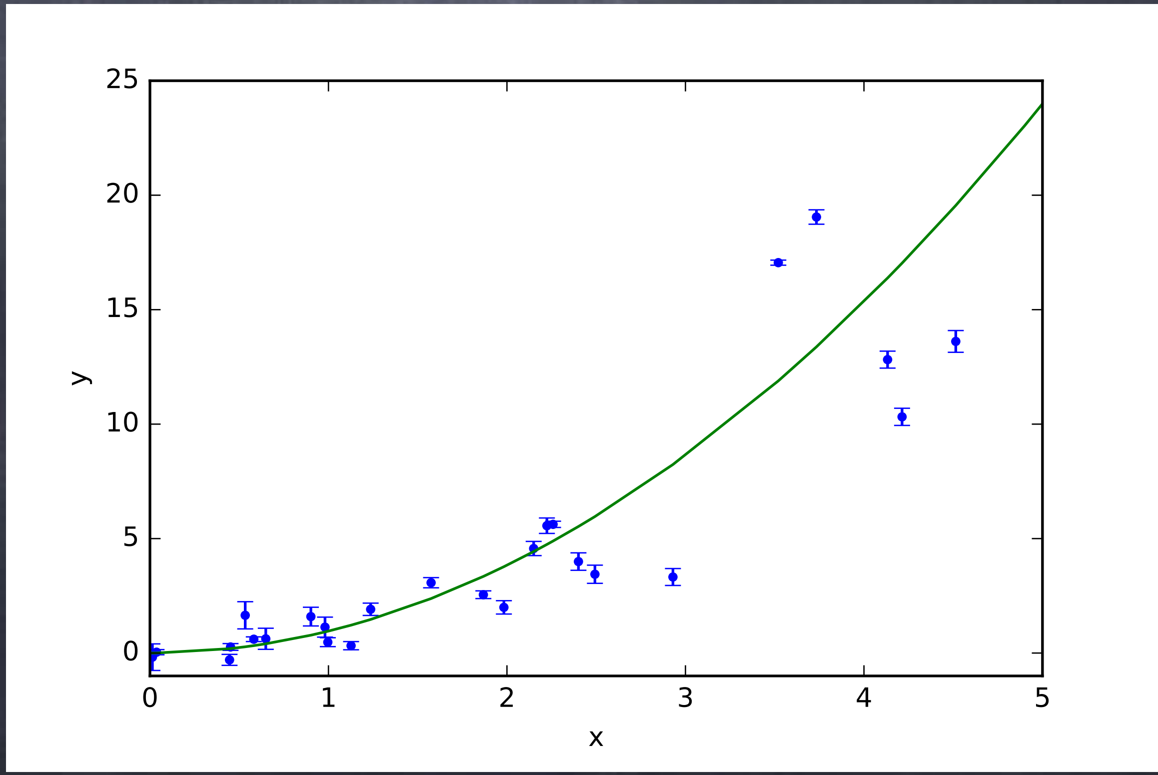
Parameter Estimation

- What do we do if we want to estimate the slope and y-intercept?



- Linear least square method

- What if data is not a straight line? And/or model is not linear, and/or if we have more than two free parameters? or more important what if I dont believe too much on the error bars*



Parameter Estimation. Level 0

- Least square method.

$$
\begin{aligned}
a &= \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i\, y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \\
&= \frac{\bar{y}\left(\sum_{i=1}^{n} x_i^2\right) - \bar{x} \sum_{i=1}^{n} x_i\, y_i}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} \\
b &= \frac{n \sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \\
&= \frac{\left(\sum_{i=1}^{n} x_i\, y_i\right) - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}
\end{aligned}
$$

Where this comes from?

Minimize the residual

$$R^2 \equiv \sum \left[y_i - f(x_i, a_1, a_2, \ldots, a_n)\right]^2$$

assuming:

- A linear function f=ax+b.

- Errors are Gaussian and uncorrelated.

Minimization implies:

$$\frac{\partial R^2}{\partial a_i} = 0$$

Exercise: Write a python code that finds a and b, for a given data set. Show an example.

Parameter Estimation. Level 1

- χ^2 Minimization $\qquad \dfrac{\partial \chi^2}{\partial \theta} = 0$

$$\chi^2 = \sum (y_i - y(x_i, \theta))^2 / \sigma_{y_i}^2$$

θ : free parameters

σ_{y_i} : variance ony_i

Exercise: show explicitly that the linear least square method is derived from the mininimization of the chisquare when the model is a straight line.

Parameter Estimation and optimization

$$\chi^2$$

$$\chi^2 = \sum (y_i - y(x_i, \theta))^2 / \sigma_{y_i}^2$$

$$\frac{\partial \chi^2}{\partial \theta} = 0$$

$$\theta : \text{free parameters}$$

$$\sigma_{y_i} : \text{ variance on} y_i$$

Chisq minimization becomes difficult (sometimes imposible) when the number of parameters increases...
We can do the minimization of Chi^2 with MonteCarlo sampling.

Least square, and minimum Chi^2 methods are just special cases of Statistical Inference. This Chi^2 is a gaussian distribution if data points are independent, and errors are also Gaussian.

Probability
(some definitions)

Typical answer

- "The ratio of the number of favorable cases to the number of all cases"
- " The ratio of the number of times the event occurs in a test series to the total number of trials in the series"

A subjective definition

- A formal definition would be: "The quality, state, or degree of something being supported by evidence strong enough make it likely though not certain to be true"
- A simple definition: "A measure of the degree of belief that an event will occur"

Probability Rules

$$0 < p(x) < 1$$

Probability of event x happens is coherent

$$p(x) + p(\sim x) = 1$$

Probability that event "x" happen, and probability of event x do not happen are complementary.

$$p(x, y) = p(x|y)p(y)$$

Product rule

$$p(x) = \sum_i p(x, y_i)$$

Probability that event "x" happen, given that y happened: Marginalization

$$p(x) = \int p(x, y)dy$$

In the continuous limit we change the sum by an integral.

BAYES THEOREM arise from the these rules. Next Class.

Probability function (discrete variable)

- To each possible value of x we associate a degree of belief.
$$f(x) = p(X = x)$$
- f(x) must satisfy the Probability rules.
- Define the Cumulative distribution function ,
$$F(x_k) \equiv P(\leq x_k) = \sum_{xi \leq x_k} f(x_i)$$ CDF
- with properties: $$F(-\infty) = 0$$
$$F(\infty) = 1,$$
- Also define de mean, or expected value.

$$\mu = \bar{x} = E(x) = \sum_i x_i f(x_i)$$

In general: $$E(g(x)) = \sum_i g(x_i) f(x_i)$$ $$E(aX + b) = aE(X) + b$$

- The standard deviation and Variance .

$$\sigma^2 \equiv Var(X) = \bar{X^2} - \bar{X}^2 \qquad \sigma = \sqrt(\sigma^2)$$

$$Var(aX + b) = a^2 Var(X),$$

- The mode, or the most probable value (for unimodal functions)

$$\left(\frac{df(x)}{dx} \right)_{x_m} = 0$$

Probability density function (continus variable)

- The degree of believe of each value is quantified by the probability density function, pdf. $f(x)$

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

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x')dx' \qquad \text{CDF}$$

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

$$E(g(X)) = \int_{\infty}^{\infty} g(x)f(x)dx$$

$$\sigma^2 \equiv Var(X) = \bar{X}^2 - \bar{X}^2$$

Centra Limit Theorem

- The mean and variance of a linear combination of random variables is given by:

$$Y = \sum_{i=1}^{n} c_i X_i$$

$$\sigma_Y^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2$$

- CLT: The distribution of a linear combination Y will be approximately normal if the variables X_i are independent and \sigma_Y^2 is much larger than any single component c_i^2 \sigma_i^2 from a non-normally distributed X_i.

Two applications of CLT

- 1) A sample avergage $\bar{X}_n$ of n independent identical distributed variables ,

$$\bar{X}_n = \sum \frac{1}{n} X_i$$

is normally distributed, since it is a linear combination of n variables X_i with c_i=1/n then:

$$\bar{X}_n \sim N(\mu_{\bar{X}_n}, \sigma_{\bar{X}_n})$$

- 2) Binomial, Poisson and χ^2 distribution can be approximated, for large numbers, by a Gaussian distribution. Other

- Example: Binomial distribution $X \sim B_{n,p}$

$$f(x|B_{n,p}) = \frac{n!}{(n-x)!x!} p^x (1-p)^{(n-x)},$$

$$n = 1, 2, 3..., n$$

$$0 \leq p \leq 1$$

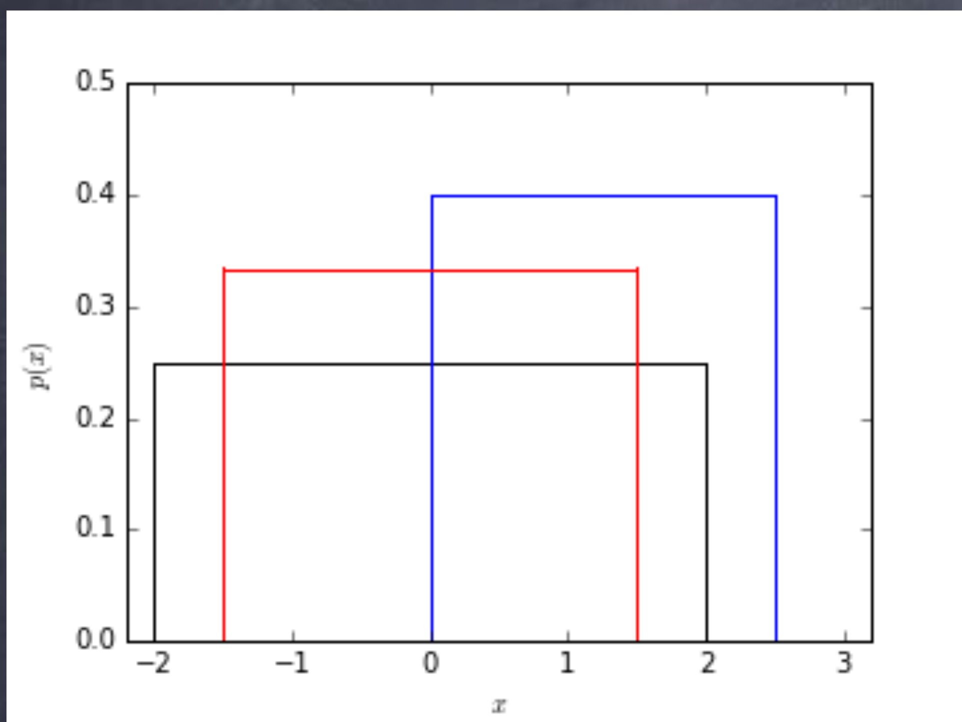
$$x = 0, 1, ..., n$$

$$\mu = np$$

$$\sigma = \sqrt{(np(1-p))}$$

Examples of probability distributions

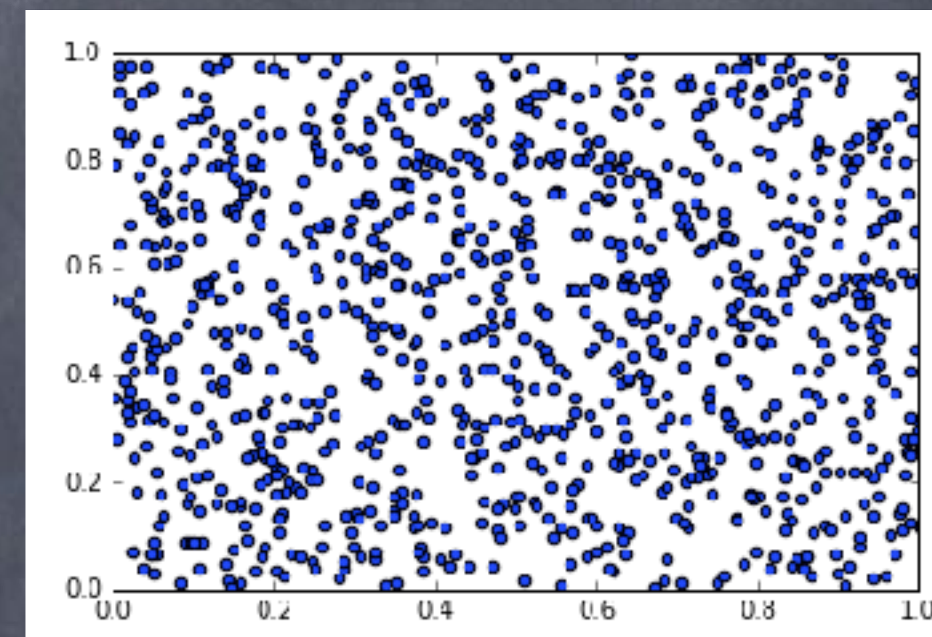
Probability distribution functions:
The basic one: Uniform



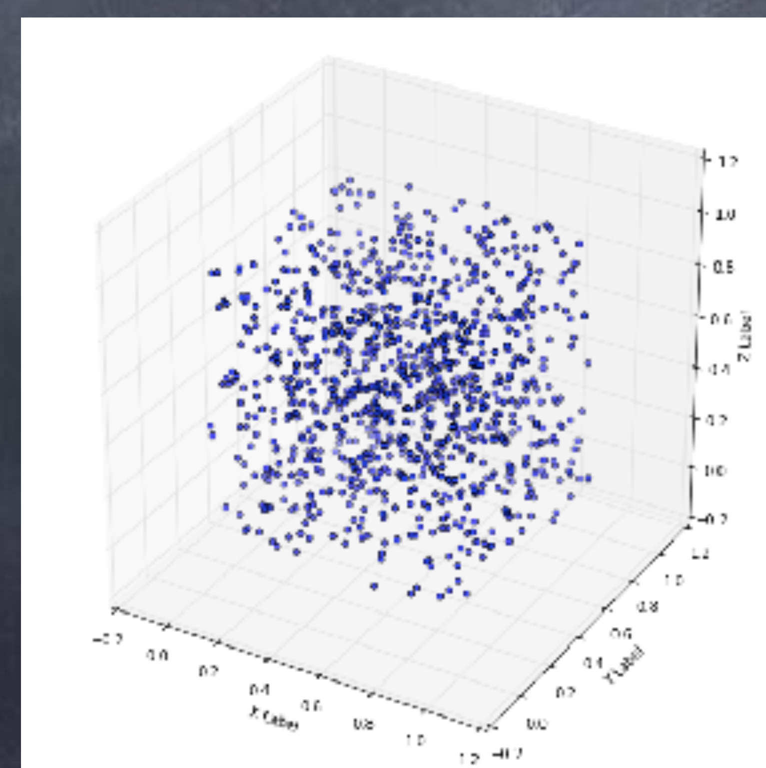
$$p(x \mid \mu, W) = \frac{1}{W} \text{for} \mid x - \mu \mid \leq \frac{W}{2}$$

$$W = b - a$$

1D



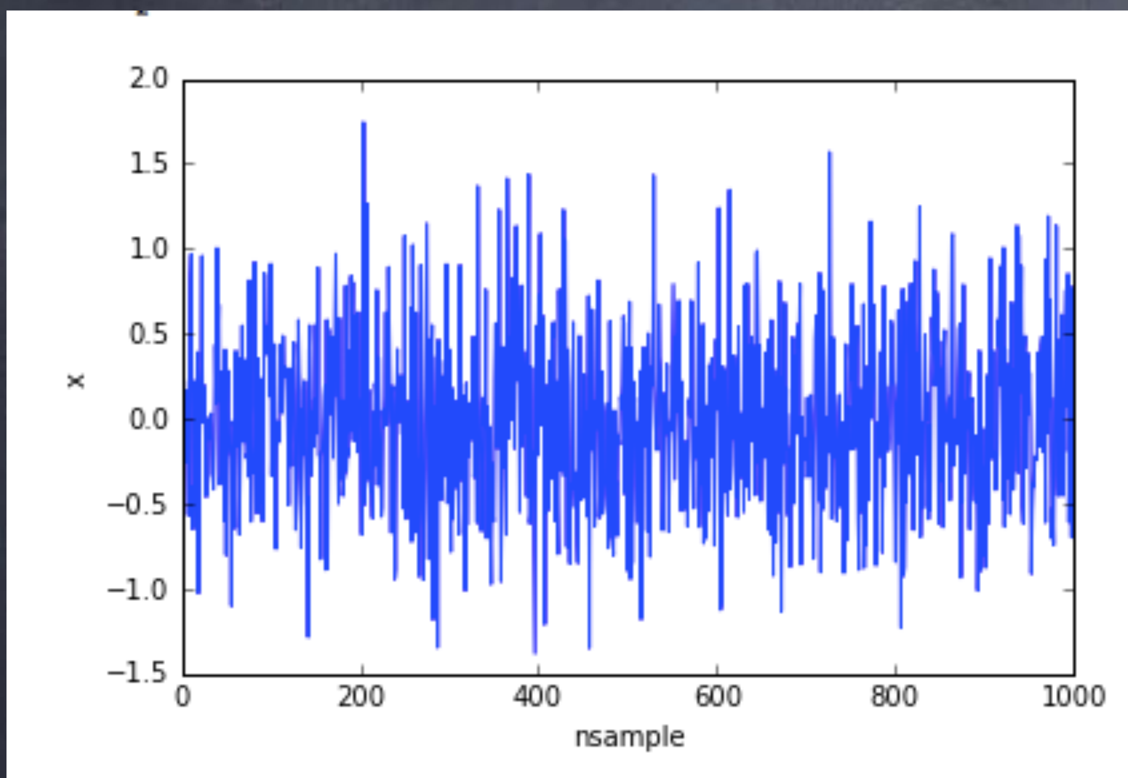
2D

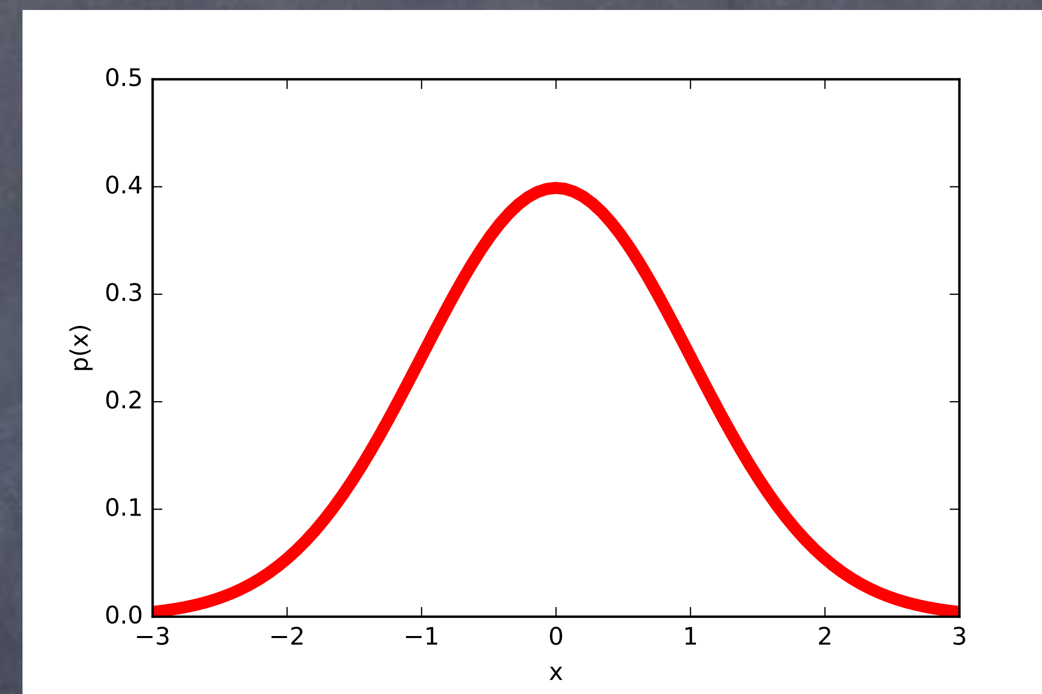


3D

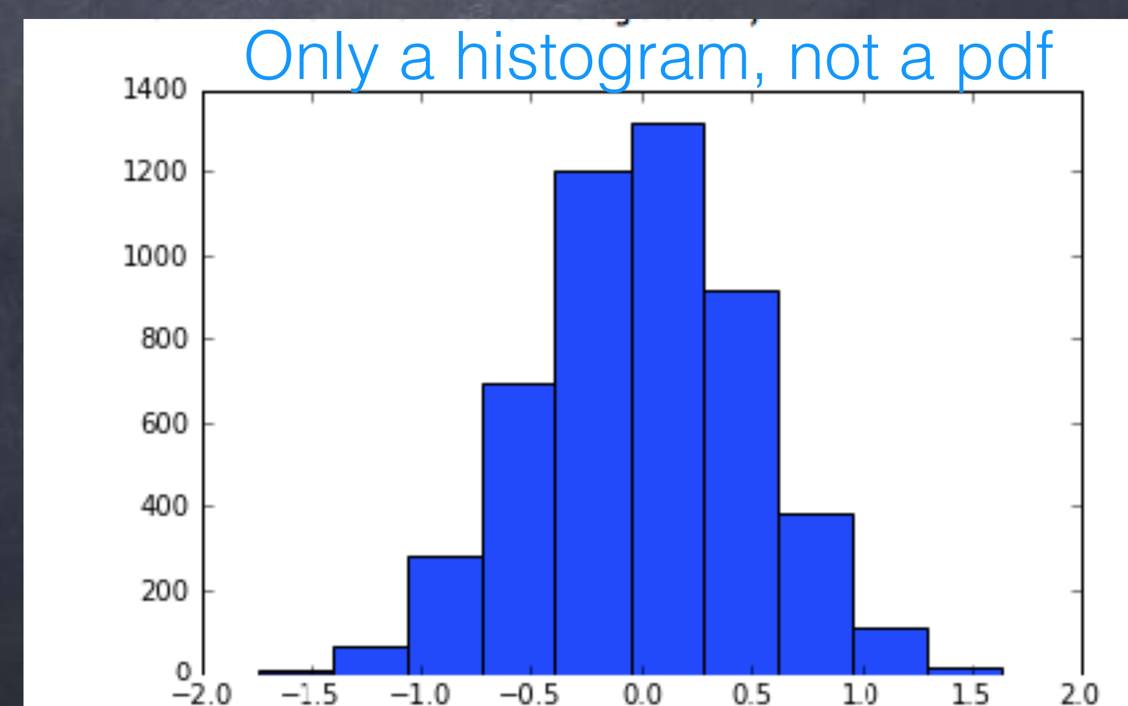
Probability distribution functions:

The next basic one: Gaussian/Normal





$$p(x\,|\,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{-(x-\mu)^2}{2\sigma^2}\right)$$



Properties

- The convolution of two gaussian distribution is gaussian.
- Ej. $\mu_c = \mu_0 + b$ and $\sigma_c = \sqrt{\sigma_0^2 + \sigma_e^2}$ Where mu_0 and sigma_0 defines the distribution of some quantity we want to measure, and b and sigma_e defines de error distribution.

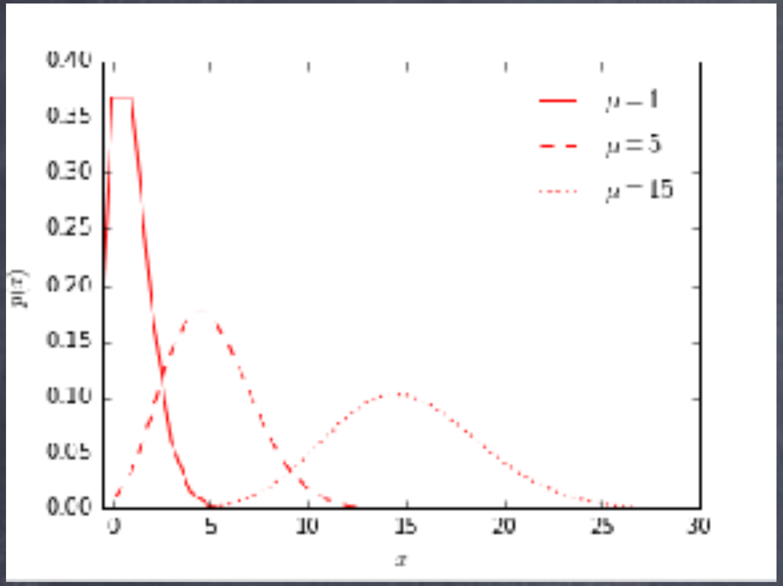
Convolution

$$(f \star g)(x) = \int_{-\infty}^{\infty} f(x')g(x - x')dx' = \int_{-\infty}^{\infty} f(x - x')g(x')dx'$$

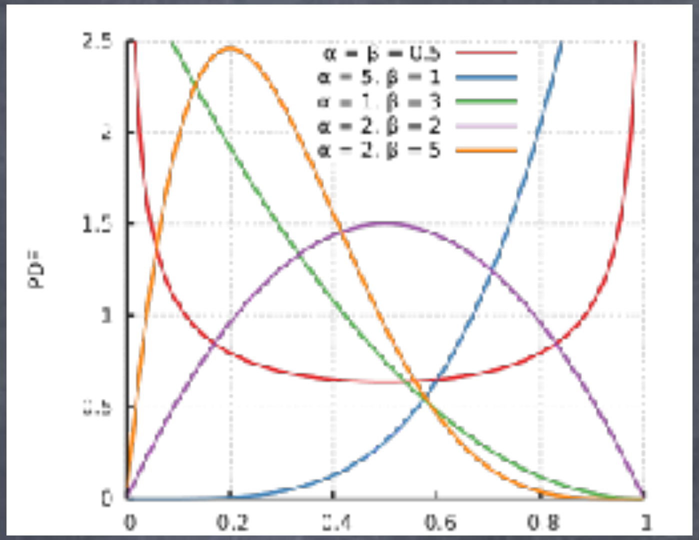
- Fourier transform of a Gaussian is a Gaussian.
- Central limit: the mean of samples drawn from almost any distribution will follow a Gaussian.

Other Probability distribution

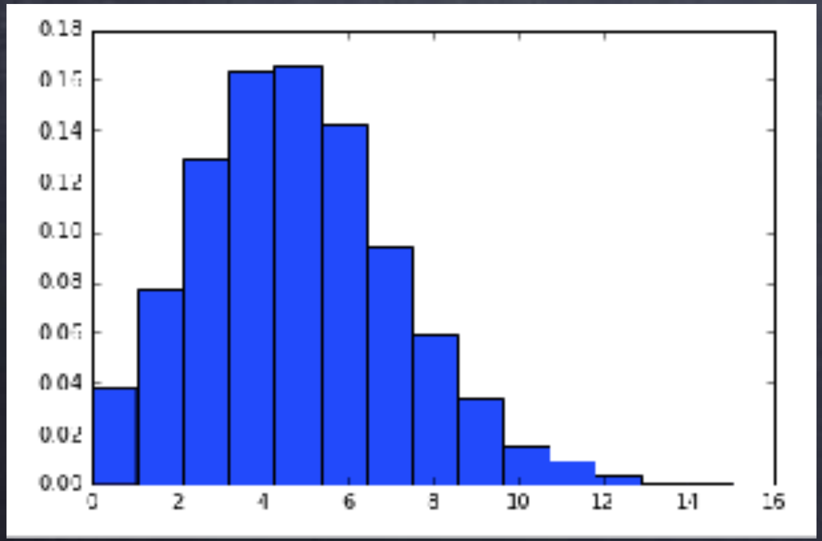
- Binomial, Poisson and distribution can be approximated, for large numbers, by a Gaussian distribution. Other distributions



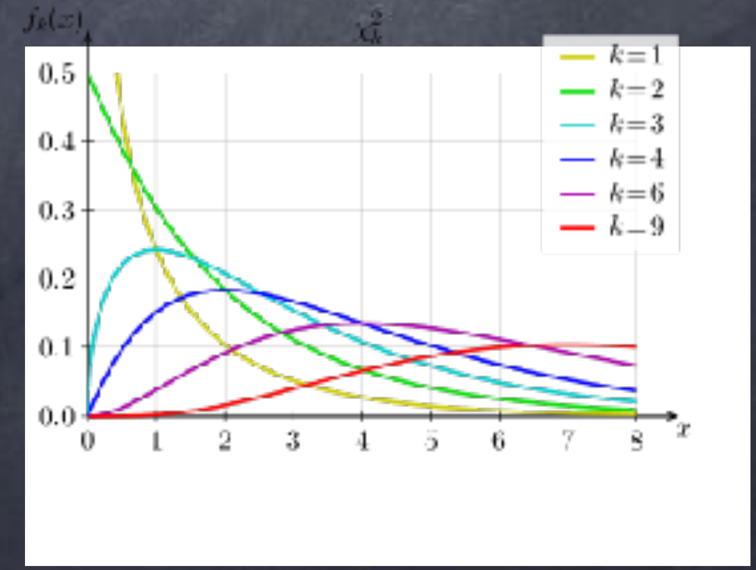
- Poisson



- Beta distribution



- Binomial



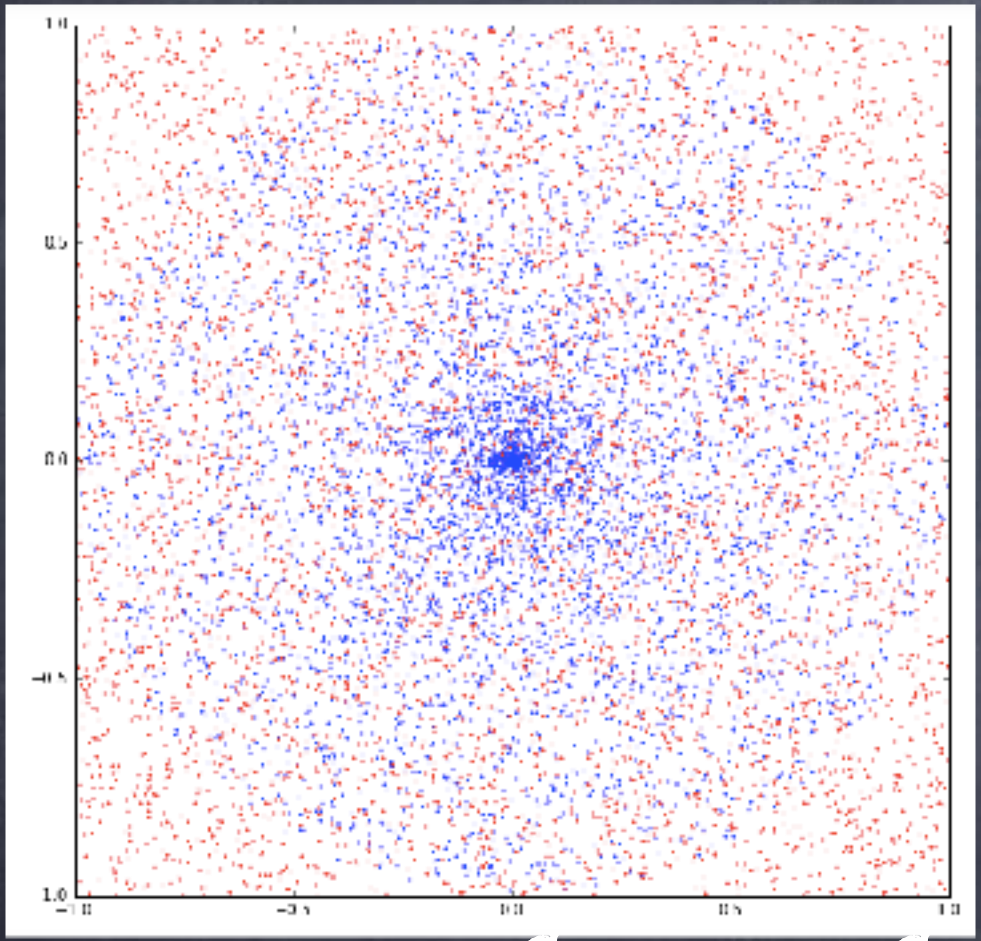
- χ^2 distribution

Exercise

- For the distributions mentioned, find the CDF, mean, median, mode, variance and standard deviation. Plot both the PDF and CDF, for some different values of mean and sigma.
- Investigate about other useful distribution functions.

Transformation of variables.

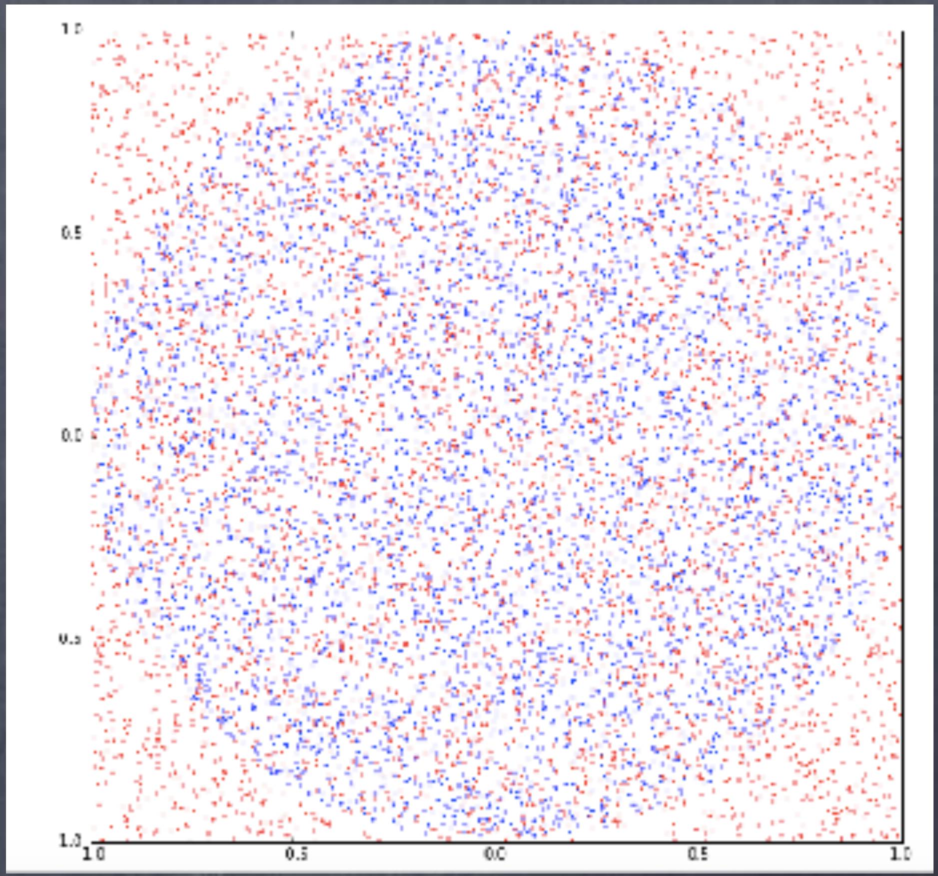
(Ej. 2D)



non-Uniform for r<1

1) generate r and theta, from U.

!=

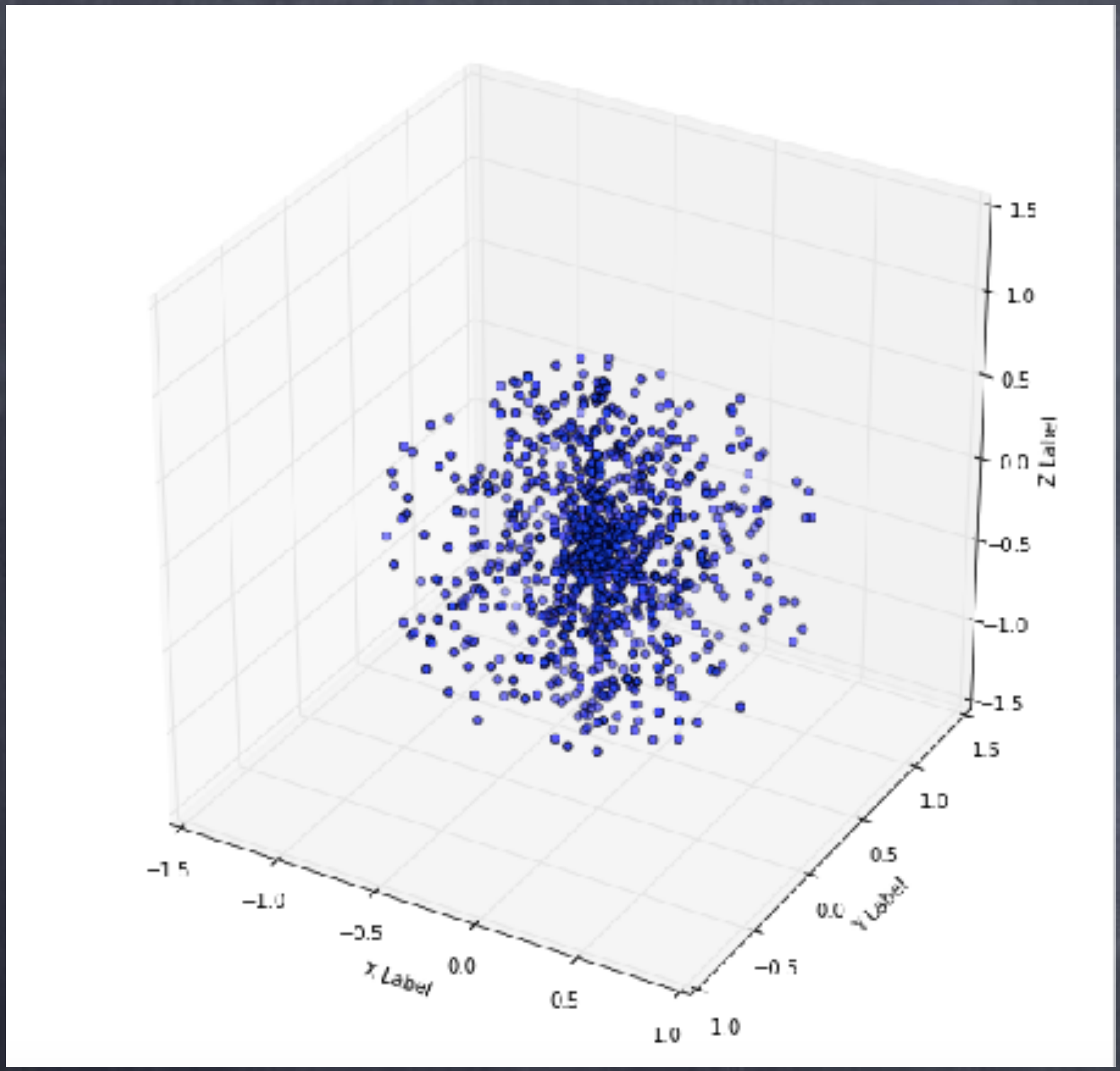


Uniform for r<1

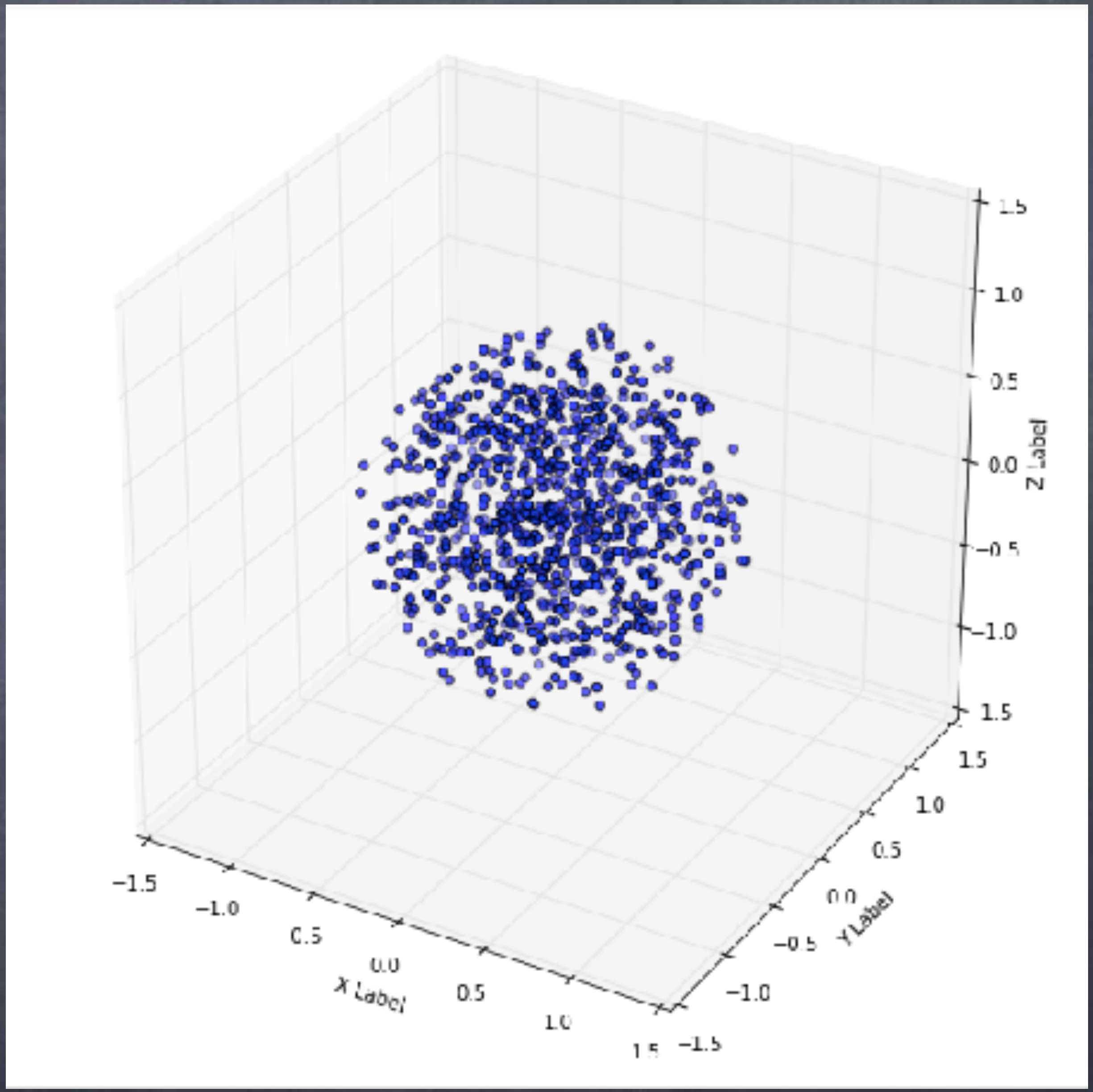
1) generate x,y from U so that r=sqry(x^2+y^2)<=1

2) generate sqrt[r] and theta, from U.

Transformation of variables.



1) generate r and theta, phi from U.



1) generate x,y from U
r=sqry(x^2+y^2+z^2)<=1
2) generate sqrt[r], theta and z=cos(phi) from U.

Transformation of variables.

- Any function of a random variable is a random variable itself.
- Sometimes we measure a variable x, but the interesting final result is y(x). In we know the PDF p(x), what is the PDF p(y)?, where y=Phi(x).

$$p(y) = p[\Phi^{-1}(y)] \left| \frac{d\Phi^{-1}(y)}{dy} \right|$$