FALL 2006

1

Fundamental Statistical Measures

Refs: Davis-notes chap. 2

Statistic for a single variable (Moments)

Def: X a random variable whos values are defined by a random process, a process which produces values not perfectly known (e.g. the Lorenz attractor example)

Many realization of x produce an ensemble x_n

The "expected value" or "mean" or "average"

$$\langle x \rangle = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} x_n$$

if N is finite it becomes a sample estimate and may lead to a "bias" estimate of the mean.

The expected value has certain properties:

$$\langle x + y \rangle = \langle x \rangle + \langle y \rangle$$

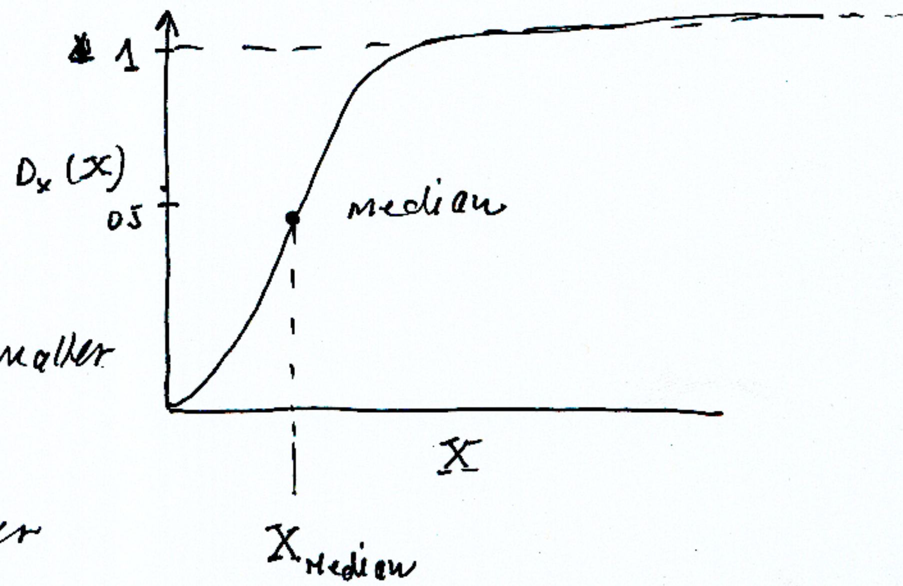
$$\langle a x \rangle = \langle x a \rangle = a \langle x \rangle \quad \text{if } a = \text{constant}$$

$$\langle x y \rangle \neq \langle x \rangle \langle y \rangle$$

DISTRIBUTION FUNCTION and PDF

A complete description of x is given by the distribution function $D_x(X)$

$$D_x(X) = \text{the fraction of } x < X$$



X_{Median} is the value of x for which 50% of the x are smaller than X_{Median} and the remaining 50% is bigger

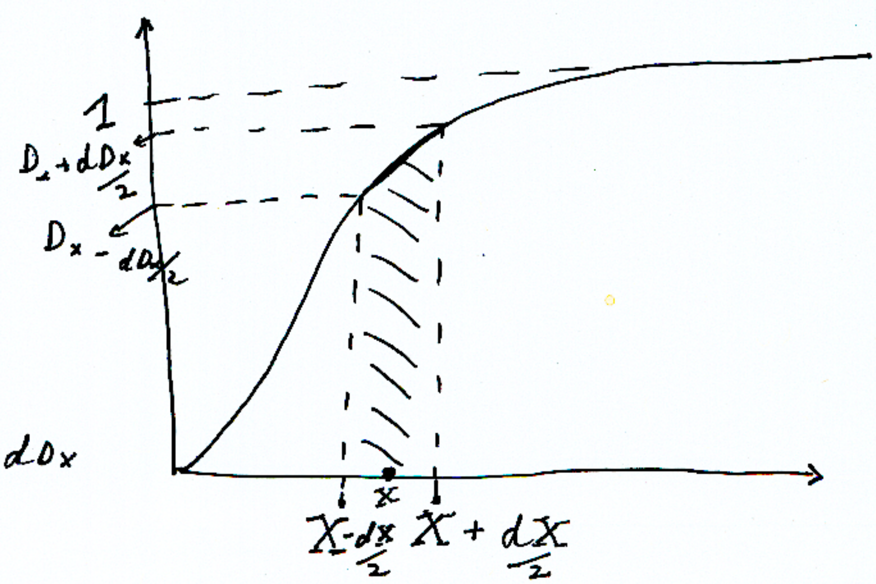
$$D_x(+\infty) = 1$$

$$D_x(-\infty) = 0$$

"The Probability Distribution Function" (PDF)

what is the fraction of occurrences x falling in the interval

$$X - \frac{dX}{2} < x < X + \frac{dX}{2}$$



$\cancel{= D_x + dD_x - D_x + dD_x}$

$$= \left(D_x + \frac{dD_x}{2}\right) - \left(D_x - \frac{dD_x}{2}\right) = dD_x$$

~~The PDF is the probability that any given~~

The probability that an x will fall in that interval is given by the slope of the distribution function $D_x(X)$. For $dX \to 0$

The slope is given by:

$$\lim_{dX \to 0} \frac{\left(D_x + \frac{dD_x}{2}\right) - \left(D_x - \frac{dD_x}{2}\right)}{\left(X + \frac{dX}{2}\right) - \left(X - \frac{dX}{2}\right)} = \frac{d\,D_x(X)}{dX} \equiv P_x(X)$$

derivative of distribution function

PDF

$P_x(X) \equiv$ "Probability Distribution Function", it is the probability that a given $x = X$ (the value X)

NOTE:

1) $$P_x(X)\,dX = \frac{d\,D_x}{dX}\,dX = d\,D_x$$

mutiplying the pdf by dX gives you the fraction of occurrences of x falling in the interval $dX = X_2 - X_1$

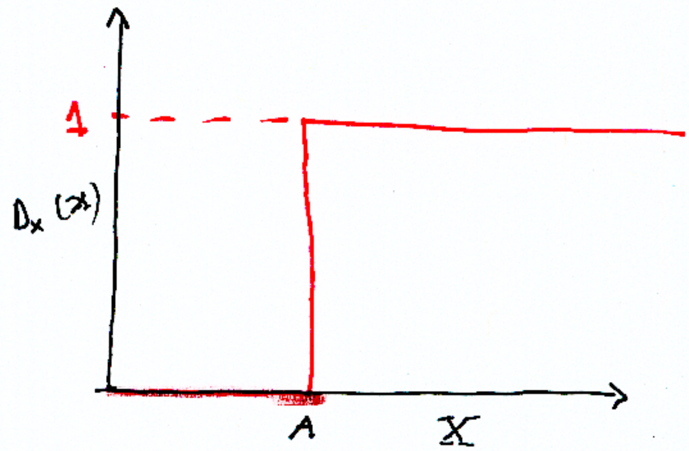
2) The integral of

$$\int_{-\infty}^{\infty} P_x(X)\,dX = \int_{-\infty}^{\infty} \frac{d\,D_x}{dX}\,dX = \Big[D_x\Big]_{-\infty}^{+\infty} = \overset{1}{D_x(\infty)} - \overset{\emptyset}{D_x(-\infty)}$$

$$= \underline{\underline{1}}$$

Example:

Assume the following distribution function.



This is called a Heavyside function. It means that all the values occur when $x = A$.

$$P_x(x) = \delta(x - A) \quad \begin{cases} = 1 & \text{for } x = A \\ = 0 & \text{for } x \neq A \end{cases}$$

recall $P_x(x) = \dfrac{dD_x}{dx} \Rightarrow D_x(x) = \displaystyle\int_{-\infty}^{x} P_x(x')\,dx' =$

$$= \int_{-\infty}^{x} \delta(x' - A)\,dx'$$

Heavyside function.

HANDY NOTATION FOR PDF

1) Assume a large number N of realizations of x_n

~~2) Assume that # of the x_n are xxxxx~~

$x_1, x_2, x_3 \ldots x_N$

$$P_x(x) = \frac{1}{N}\left[\delta(x - x_1) + \delta(x - x_2) + \ldots \delta(x - x_n)\right]$$

in the limit $N \to \infty$

$$P(x) = \lim_{N \to \infty} \frac{1}{N} \sum_1 \delta(x - x_n) = \langle \delta(x - x) \rangle$$ ← the expected value

It is now instructive to compute the

MOMENTS of x

$$\int_{-\infty}^{\infty} P_x(X)\, dX = \int_{-\infty}^{\infty} \langle \underbrace{\delta(X - x)}_{\substack{\text{is one when} \\ X = x \\ \text{and } \emptyset \text{ otherwise}}} \rangle\, dX = \langle 1 \rangle = 1$$

1st moment

$$\mu_1 = \int_{-\infty}^{\infty} X P_x(X)\, dX = \langle x \rangle$$ = expected value or average of x

2n moment

$$\mu_2 = \int_{-\infty}^{\infty} X^2 P_x(X)\, dX = \langle x^2 \rangle$$

n^{th} moment

$$\mu_n = \int_{-\infty}^{\infty} X^n P_x(X)\, dX = \langle x^n \rangle$$

for $n > 1$ usually the random variable x' is defined as an anomaly with respect to the expected value

$$x' = x - \langle x \rangle$$

$\mu_2 \equiv$ variance = energy of the process

$\sigma = \sqrt{\mu_2}$ = standard deviation = typical fluctuation

- knowledge of all the moments is sufficient to characterize the entire PDF.

FUNCTION OF A RANDOM VARIABLE

Any function of a random variable is also a random variable and its statistics can be determined by the original variable

suppose $y = g(x) = \int_{-\infty}^{+\infty} g(X)\,\delta(X - x)\,dX$

if we take the expected value

$$\langle y \rangle = \langle g(x) \rangle = \int_{-\infty}^{+\infty} g(X) \langle \delta(X-x) \rangle dX =$$

$$= \int_{-\infty}^{\infty} g(X)\, P_x(X)\, dX$$

$$= \int_{-\infty}^{\infty} Y\, P_y(Y)\, dY$$

$$\boxed{\text{now} \quad dY = dg(x) = \frac{dg}{dx}\,dx}$$

$$\int_{-\infty}^{\infty} Y P_y(Y)\, dY = \int_{-\infty}^{\infty} Y \underbrace{P_y(Y) \frac{dg}{dx}}\, dX = \int_{-\infty}^{\infty} g(X) \underbrace{P_x(x)}\, dX$$

equivalent statements

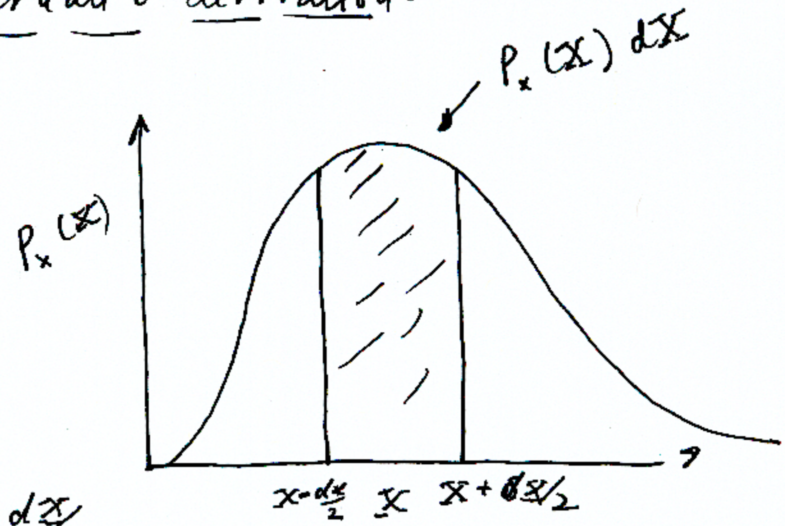
$$\boxed{P_y(Y) \frac{dg}{dX} = P_x(X)}$$

$$\boxed{P_y(Y) \frac{dY}{dX} = P_x(X)}$$

$\longrightarrow$ change of variable!

An alternative derivation.

$y = g(x)$

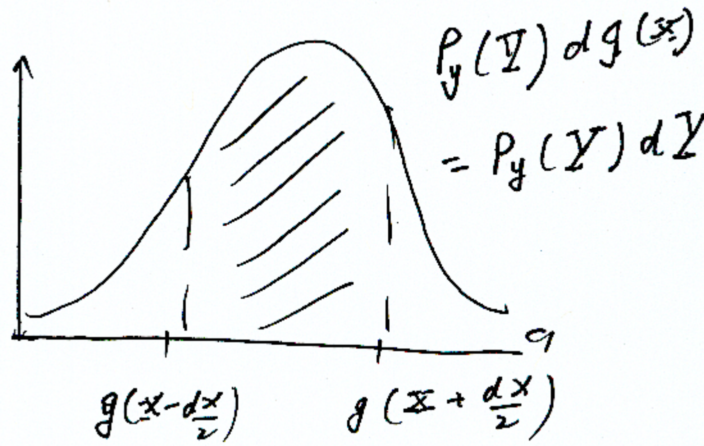


The total probability of x falling between

$X - dX/2 < x < X + dX/2$

is equivalent to the probability of $g(x)$ falling in the interval

$$g\left(X - \frac{dX}{2}\right) < g(x) < g\left(X + \frac{dX}{2}\right)$$



$$\Rightarrow P_x(X)\,dX = P_y(Y)\,dY$$

EXAMPLE 1:

Assume $y = 2x$ and $P_x(X) = \underbrace{\frac{1}{\sqrt{2\pi\mu_2}} e^{-\frac{X^2}{2\mu_2}}}_{\text{Gaussian}}$

Find $P_y(Y)$

$$P_y(Y)\, dY = P_x\, dX$$

$$P_y = P_x \left[\frac{dY}{dX}\right]^{-1}$$

$$\frac{dY}{dX} = 2$$

$$P_y = \frac{1}{2} \frac{1}{\sqrt{2\pi\mu_2}} e^{-\frac{(Y/2)^2}{2\mu_2}} = \frac{1}{\sqrt{2\pi (4\mu_2)}} e^{-\frac{Y^2}{2(4\mu_2)}}$$

same gaussian distribution with 4 x larger variance or with double standard deviation.

EXAMPLE 2:

$x = \ln y$ with x gaussian

$$P_y = P_x \frac{dx}{dy} = \frac{1}{y} P_x = \underbrace{\frac{1}{y} \frac{1}{\sqrt{2\pi \mu_2}} e^{-\frac{(\ln y)^2}{2\mu_2}}}_{\text{log normal distribution}}$$

Other PDFs

Gaussian $p_x = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$E[X]$ = 1^{st} moment expected value $= \int_{-\infty}^{+\infty} X p_x \, dx = \mu$

$Var[X]$ = 2n moment ~~mean~~ of $X' = X - \mu$ $= \sigma^2 = \int_{-\infty}^{\infty} (X-\mu)\, p_x (x-\mu)\, dX$

Lognormal, Gamma, Exponential, Chi-square, Pearson III, Beta, Gumbel, Weibull, Mixed Exponential....