# Properties of probabilities, prob. distr. functions

- My notation:

$$p(x) = \begin{cases} \text{probability (mass) for } X & \text{Units: } [p(x)] = 1 \\ \text{or} & \\ \text{probability \underline{density} at } X & \text{Units: } [p(x)] = \dfrac{1}{[x]} \end{cases}$$

- When multiple variables:

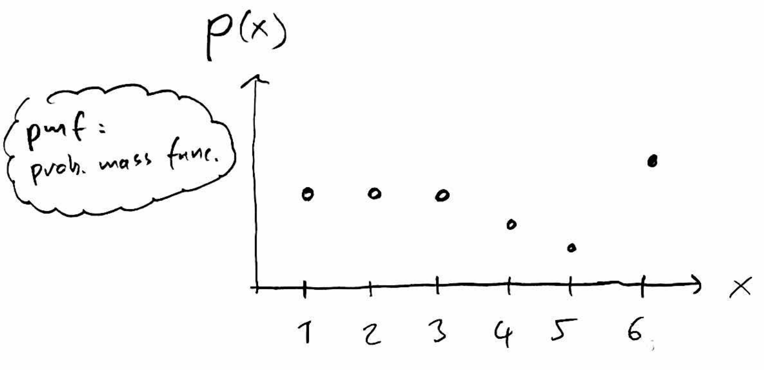
- Should do: $P_x(x)$, $P_y(y)$, $P_{x,y}(x,y)$, $P_{x|y}(x|y)$

or alternatively: $f(x)$, $g(y)$, $h(x,y)$, $q(x)$

- But I will be a bit sloppy: $p(x)$, $p(y)$, $p(x,y)$, $p(x|y)$

- Domains and probabilities:

Discrete case



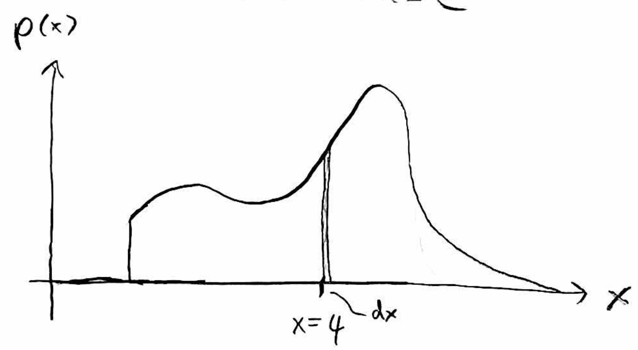

$\text{Prob}(x=4) = p(4)$

$\text{Prob}(x \le 4) = p(1)+p(2)+p(3)+p(4)$

$\text{Prob}(2 \le x \le 4) = p(2)+p(3)+p(4)$

Domain: $x \in \{1,2,3,4,5,6\}$

Continuous case




$\text{Prob}(x \in [4, 4+dx]) = p(4)\,dx$

$\text{Prob}(x \le 4) = \int_{-\infty}^{4} p(x)\,dx$

$\text{Prob}(2 \le x \le 4) = \int_{2}^{4} p(x)\,dx$

Domain: $x \in \mathbb{R}$.

- $X$ is a stochastic variable (random variable)

- We say that $X$ "has a pdf $p(x)$", or "follows a pdf $p(x)$", or "is distributed as $p(x)$", etc.

- Shorthand (but potentially confusing) notation:

$$X \sim p(x)$$

[Does not mean that "$x$ is approximately equal to $p(x)$" or that "$x$ is proportional to $p(x)$"!]

- A function of random variables is itself a random variable

Example: Throw two dice

$X_1$: Outcome for die 1 $\quad p_{X_1}(x_1) = \frac{1}{6} \quad x_1 \in \{1,2,3,4,5,6\}$

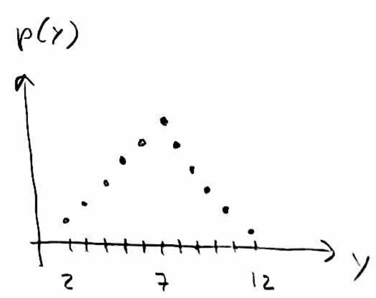
$X_2$: Outcome for die 2 $\quad p_{X_2}(x_2) = \frac{1}{6} \quad x_2 \in \{1,2,3,4,5,6\}$

Let $Y \equiv X_1 + X_2$.

Domain: $Y \in \{2, 3, \ldots, 11, 12\}$

Prob. distr.:

$$P_Y(y) = \begin{cases} 1/36 & y = 2,\ 12 \\ 2/36 & y = 3,\ 11 \\ 3/36 & y = 4,\ 10 \\ 4/36 & y = 5,\ 9 \\ 5/36 & y = 6,\ 8 \\ 6/36 & y = 7 \end{cases}$$



× end lecture

- Probabilities are numbers in [0,1]

Discrete: $0 \le p(x) \le 1$

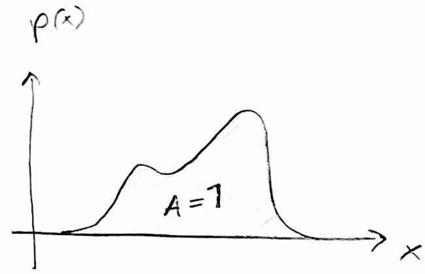
Cont.: $0 \le p(x)dx \le 1$

Note: prob. density $p(x)$ can have arbitrarily large numerical value (depends on choice of units for $x$), but must always be positive.

- Pdfs are normalized to unity:

Discrete: $$\sum_{x \in D} p(x) = 1$$

Cont.: $$\int_{x \in D} p(x)\,dx = 1$$

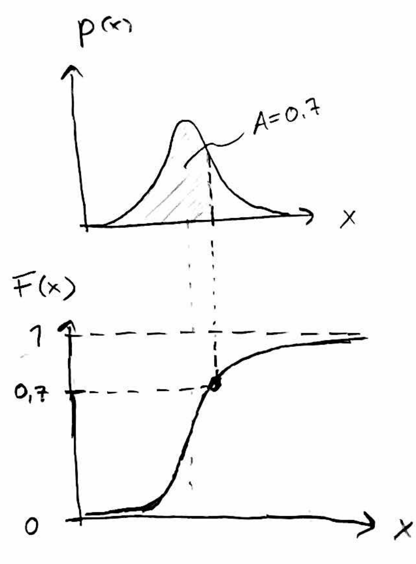



- We'll also need the cumulative prob. distribution function (cdf)

$$F(x) \equiv \mathrm{Prob}(X \le x) = \int_{-\infty}^{x} p(x')\,dx'$$

- Relation to pdf:

$$p(x) = \frac{d}{dx} F(x)$$



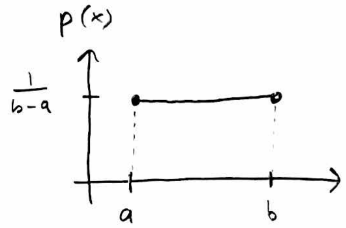
# Some important (1D) prob. distributions

$$\theta(y) = \begin{cases} 0 & \text{for } y < 0 \\ 1 & \text{for } y \geq 0 \end{cases}$$

- The <u>uniform</u> distribution

$$p(x) = \frac{1}{b-a}\,\theta(x-a)\,\theta(b-x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

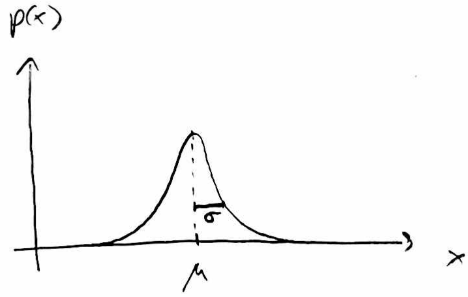
- Parameters: $a, b$ $\quad [a]=[b]=[x]$
- standard form: $a=0$, $b=1$




- The <u>Gaussian / normal</u> distribution

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in (-\infty, \infty)$$

- Parameters: $\mu, \sigma$ $\quad [\mu]=[\sigma]=[x]$
- Standard form: $\mu=0$ (location), $\sigma=1$ (scale)



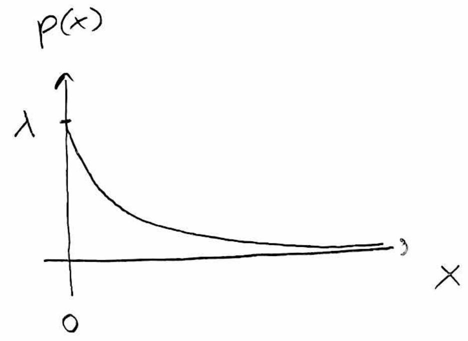

{ Pops up everywhere because of the Central Limit Theorem }

- The <u>exponential</u> distribution

$$p(x) = \lambda e^{-\lambda x} \quad , \quad x \in [0, \infty)$$

- Parameters : $\lambda$

Note : $[\lambda] = \frac{1}{[x]}$




[Note : The Boltzmann distr. is an exp. distr.
(Project 4)]

# A couple of discrete prob. functions

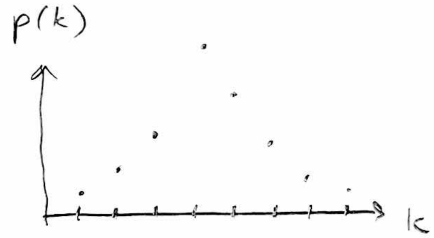
- The binomial distribution

$$p(k) = \binom{n}{k} p_{succ.}^{k} (1 - p_{succ})^{n-k}$$

$$\binom{n}{k} = \text{binomial coeff} = \frac{n!}{k!(n-k)!}$$

$p(k) = \text{Prob}\left(\begin{array}{l}\text{get exactly } k \text{ successes in} \\ n \text{ indep. Bernoulli (yes/no) trails}\end{array}\right)$

- Parameters: $n$ (num. of trails)
  $p_{succ.}$ (prob. of success in a trail)

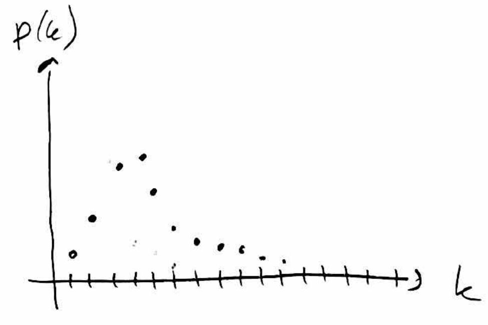



- The Poisson distr.

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

$p(k) = \text{Prob}\left(\begin{array}{l}\text{get } k \text{ events in some interval when events} \\ \text{occur at some known mean rate } \lambda\end{array}\right)$

- Parameters: $\lambda \in [0, \infty)$ (rate)




Poisson distr. is the limit of the binomial when
$n \to \infty$
$p_{succ} \to 0$
in such a way that the product
$n p_{succ} \to \lambda$

LHC physics!

# Prob. dens. functions of many variables

- Notation : $p(x_1, x_2, x_3, \dots)$ or $p(\bar{x})$

  or $p(x,y)$ in the case of two variables

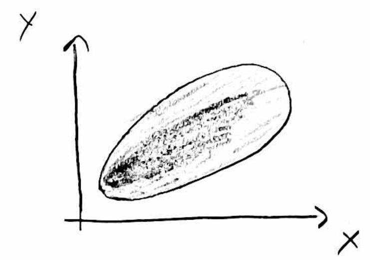
- Look at 2D examples :

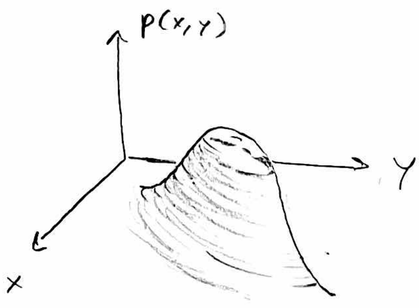
- Need to distinguish

  - joint prob. dens. $p(x,y)$
  - conditional prob. dens. $p(x|y)$ , $p(y|x)$
  - marginal prob. dens. $p(x)$ , $p(y)$

- Joint pdf :

  - $p(x,y)\,dx\,dy = \text{Prob}\left(X \in [x, x+dx] \text{ and } Y \in [y, y+dy]\right)$
  - Prob. dens. over a 2D space (xy-plane)



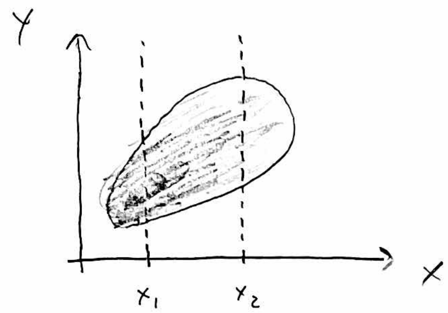




- Normalization : $\iint p(x,y)\,dx\,dy = 1$

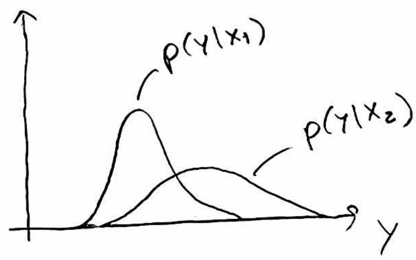
- Conditional pdfs

  - $p(Y|X)\,dy$ = Prob ( $Y \in [y, y+dy]$ given a specific $X=x$ )

    [analogous for $p(x|y)$ ]
  - Prob dens. over a 1D space (in this example : Y axis)
  - Example: If the joint pdf $p(x,y)$ looks like this ...




...we can get conditional pdfs looking like this

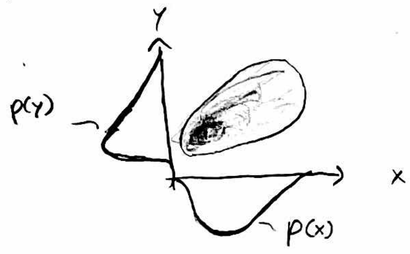



- Marginal pdfs

  - $p(x)\,dx$ = Prob ( $X \in [x, x+dx]$ , independent of Y )
  - Prob dens. over a 1D space (here: x axis)   [analogous for $p(y)$ ]

$$p(x) = \int p(x,y)\,dy$$ "marginalise over y"

$$p(y) = \int p(x,y)\,dx$$ " —— " —— x "



- Useful relations:

Bayes theorem $\quad P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

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

2) Discrete: $$p(x) = \sum_{y \in D} p(x,y) = \sum_{y \in D} p(x|y)p(y)$$

$$p(y) = \sum_{x \in D} p(x,y) = \sum_{x \in D} p(y|x)p(x)$$

Continuous: $$p(x) = \int p(x,y)\,dy = \int p(x|y)p(y)\,dy$$

$$p(y) = \int p(x,y)\,dx = \int p(y|x)p(x)\,dx$$

The conditional pdfs weighted according to other marginal pdf.

With 1) and 2) we can write Bayes theorem as

Discrete: $$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\sum_{y \in D} p(x|y)p(y)}$$

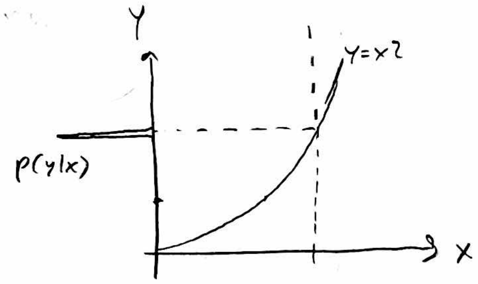
Cont.: $$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\int p(x|y)p(y)\,dy}$$

- Sometimes a "deltafunction perspective" is useful

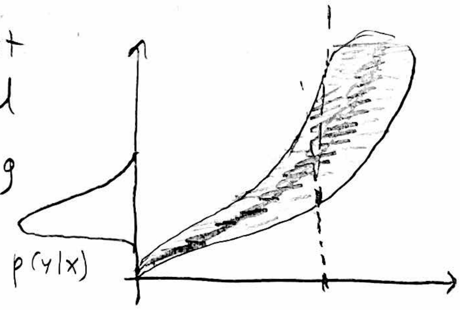
<u>Instead of</u>:
- $X$ is rand. variable with pdf $p(x)$
- $Y = x^2$ is a function of $x$

<u>Rather</u>:
- $x$ and $y$ are random variables
- The statement $y = x^2$ is just saying that we are 100% certain about the <u>conditional</u> pdf $p(y|x)$, i.e.

$$p(y|x) = \delta(y - x^2) \quad \text{deltafunction!}$$




is a limit of a general case, e.g




$$p(x, y) = p(y|x)\,p(x)$$
$$= \delta(y - x^2)\,p(x)$$

$$\boxed{\begin{aligned} p(y) &= \int p(x,y)\,dx = \int p(y|x)\,p(x)\,dx \\ p_Y(y) &= \int \delta(y - x^2)\,p_X(x)\,dx \\ p_Y(y) &= p_X(X = \sqrt{y}) \end{aligned}}$$

End lect. Oct 22

X

# Expectation values

- Let $Y$ be some function of $x$, and $x \sim p(x)$

Then the expectation value of $Y(x)$ is

$$\boxed{E[Y] \equiv \int Y(x)\, p(x)\, dx}$$

some arbitrary function

A pdf

Alt. notation:

$$\langle Y \rangle \equiv \int Y(x)\, p(x)\, dx$$

We will use this notation in Project 4 because it is common in the physics literature

Discrete case:

$$E[Y] = \sum_{x \in \mathbb{D}} Y(x)\, p(x)$$

sum over possible outcomes

- So it's a weighted sum of all possible outcomes for $Y(x)$ (by running through all possible inputs $x$), where each term is weighted by the corresponding prob. $p(x)$ for the input $x$.

- Example: $Y$: money earned in a single coin flip bet:

$$Y(x) = \begin{cases} -1 & \text{for } x = \text{heads} \\ +1 & \text{for } x = \text{tails} \end{cases}$$

$$p(x) = \begin{cases} 0.5 & x = \text{heads} \\ 0.5 & x = \text{tails} \end{cases}$$

Note: Here there is no possible outcome that corresp. to the expected value

$$\Rightarrow E[Y] = Y(\text{heads})\, p(\text{heads}) + Y(\text{tails})\, p(\text{tails})$$

$$= -1 \times 0.5 + 1 \times 0.5 = \underline{\underline{0}}$$

# Moments: particularly useful expectation values

- Analogous concept in physics: $r^n \times \begin{bmatrix} \text{physical} \\ \text{quantity} \\ \text{at } r \end{bmatrix}$

$$\int r^n \times \begin{bmatrix} \text{physical quantity} \\ \text{density} \\ \text{at } r \end{bmatrix} dr$$

- Example: Given a mass density $\rho(\vec{r})$ such that

$$M = \int \rho(\vec{r})\, d^3\vec{r}$$

  - 1st moment: Center of mass $\vec{R} = \frac{1}{M}\int \vec{r}\,\rho(\vec{r})\,d\vec{r}$
    $(n=1)$

    where's the center?

  - 2nd moment: Moment of inertia $I = \int \vec{r}^2 \rho(\vec{r})\, d\vec{r}$

    How spread out is the mass?

---

- Moments of prob. distributions:

$$E[x^n] \equiv \int x^n p(x)\, dx$$

- Zeroth moment is just norm. cond.:

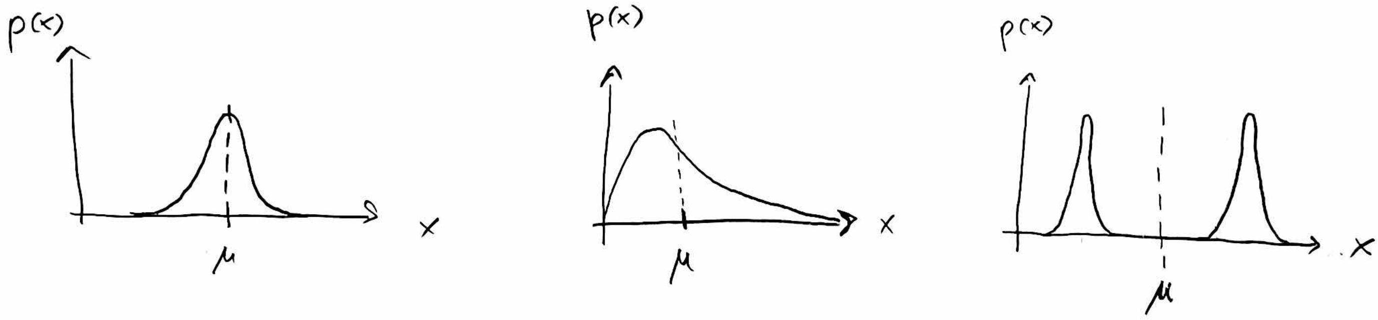
$$E[x^0] = E[1] = 1 = \int x^0 p(x)\, dx = \int p(x)\, dx$$

- First moment $E[x]$ (or $\langle x \rangle$) is called the mean of $p(x)$

$$\boxed{E[x] = \mu \equiv \int x\, p(x)\, dx}$$

The average $x$ value. (Sums $x$-values weighted by their prob.)

- Examples of means




- Moments are relative to some point

$$\int (x-c)^n \, p(x) \, dx$$

with c some constant.

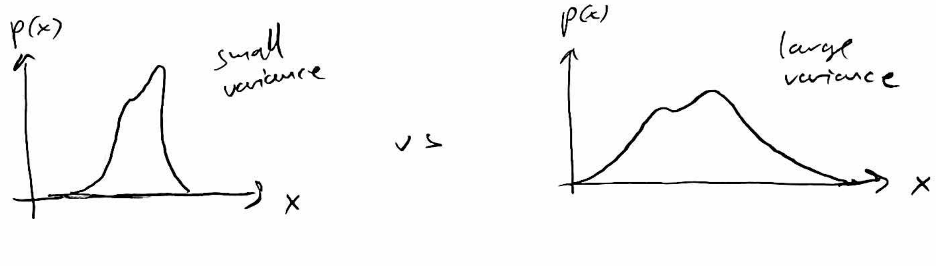
- Special case: take $c = \mu$, "central moments"

$$\int (x-\mu)^n \, p(x) \, dx$$

- $n = 0 \Rightarrow E\left[(x-\mu)^0\right] = 1$
- $n = 1 \Rightarrow E\left[(x-\mu)\right] = E[x] - \mu = \mu - \mu = 0$
- <u>$n = 2$</u> : <u>Variance</u> $(\sigma^2)$

$$Var(x) = \sigma^2 = E\left[(x-\mu)^2\right] = \int (x-\mu)^2 \, p(x) \, dx$$

A measure of how spread-out $p(x)$ is. I.e. if large values of $(x-\mu)^2$ take up much of the prob. in $p(x)$, the distribution must be <u>wide</u> $\Longleftrightarrow$ large variance.



- Useful relation

$$\begin{aligned}
Var(x) &= E[(x-\mu)^2] \\
&= \int (x-\mu)^2 p(x)\, dx \\
&= \int (x^2 - 2x\mu + \mu^2)\, p(x)\, dx \\
&= E[x^2] - 2\mu E[x] + \mu^2 \\
&= E[x^2] - 2\mu^2 + \mu^2 \\
&= E[x^2] - \mu^2 \\
&= E[x^2] - (E[x])^2
\end{aligned}$$

$$\boxed{Var(x) = E[x^2] - E[x]^2}$$

- Note that variance has units of $[x^2]$
- Standard deviation : $\boxed{\sigma = \sqrt{Var(x)}}$

[Computation of different means and variances will be important in Proj. 4.]

# Summarizing a prob. distribution with a single number

- In project 4 we will look at a lot of mean values, e.g $\langle E \rangle$

- But keep in mind that there are other <u>point estimates</u> that may be useful to summarize a pdf, e.g. the <u>median</u> and the <u>mode</u>

  - <u>median</u>: The point $x$ such that $\text{Prob}(X \leq x) = \frac{1}{2}$, i.e.

$$\int_{-\infty}^{x} p(x')dx' = \frac{1}{2}$$

  - <u>mode</u>: The point $x$ where $p(x)$ has its maximum

- For a normal distribution, the <u>mean</u>, <u>median</u> and <u>mode</u> are the same point:


Moral:
Keep in mind that $\langle x \rangle$ is not the full story. $p(x)$ is the full story!

But it's <u>not</u> true in general: