Probability

- [Discuss coin flip or dice throw]
- What does the statement $P(X) = 10\%$ mean?
- We don't know! Or at least, we don't agree...
- [Show article "Interpretations of Probability", Stanford Encyclopedia of Philosophy]
- Bertrand Russel, 1929: "Probability is the most important concept in modern science, especially as nobody has the slightest notion what it means"

- Two main interpretations

  - Frequentist: $\boxed{P(X) \equiv \lim_{n \to \infty} \frac{n_x}{n}}$ Probability defined as long-run relative frequency

  - Bayesian: $P(x) \equiv$ degree of belief / knowledge that X is true

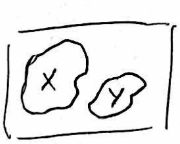
    - belief → subjective Bayesian
    - knowledge → objective Bayesian

Reasoning under uncertainty

Bruno Di Finetti: PROBABILITY DOES NOT EXIST "Theory of Prob", 1974

- Both satisfy the Kolmogorov axiom's that defines the mathematical prop. of the function $P(X)$. $\Rightarrow$ Use the same math!

Basically: $0 \leq P(x) \leq 1$



P is additive: $P(X \cup Y) = P(X) + P(Y)$ when $X \cap Y = \emptyset$

- Has several important consequences:
  - Different prob. interpretations give rise to different approaches to statistics.

    Example:
    - Bayesian statistics can ask

      $P(\text{parameter value} \mid \text{data})$?

    - Frequentist inference can not ask this question, since probability of a parameter value does not make sense. (Can only ask questions related to repeatable trails.)

  - What does <u>randomness</u> mean?
    - what's the connection between randomness and probability?
    - Is anything random (⇒ metaphysics, determinism, ...)

      ↳ <u>Apparent randomness</u> vs <u>true randomness</u>.

  - Probabilities in physics — what do they mean?

    In particular: Interpretations of quantum mechanics.

$\boxed{\text{No } \underline{\text{necessary}} \text{ link between probabilities and randomness! Can simply use prob. to express our uncertainty!}}$

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

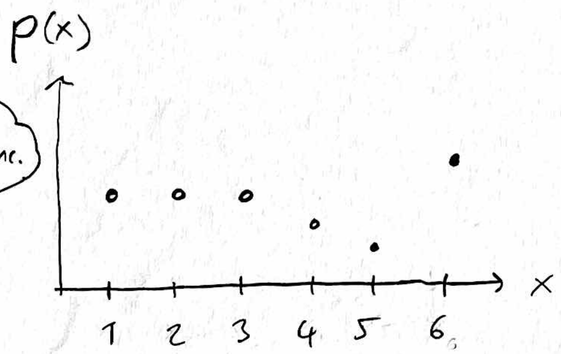
    or alternatively: $f(x)$, $g(y)$, $h(x,y)$, $q(x)$

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

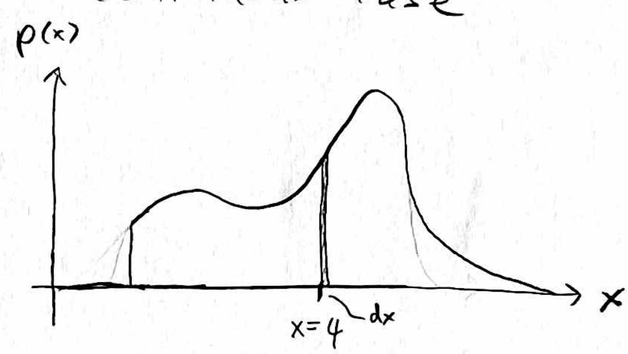
- Domains and probabilities:

Discrete case

pmf: prob. mass func.



Continuous case



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

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

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

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

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

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

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

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

- X is a <u>stochastic variable</u> (<u>random variable</u>)

- 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 <u>not</u> 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 $\quad P_{X_1}(x_1) = \frac{1}{6} \quad x_1 \in \{1,2,3,4,5,6\}$

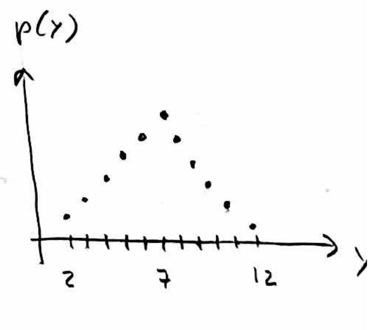
X_2 : Outcome for die 2 $\quad P_{X_2}(x_2) = \frac{1}{6} \quad x_2 \in \{1,2,3,4,5,6\}$

Let $\quad Y \equiv X_1 + X_2$

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

Prob. distr.:

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



x end lecture

- Probabilities are numbers in $[0,1]$

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

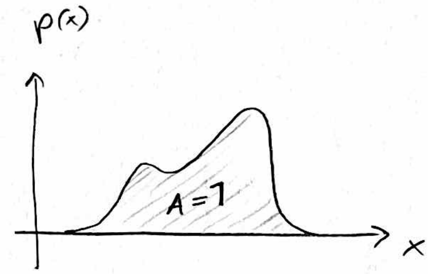
Cont.: $0 \le p(x)dx \le 1$

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:

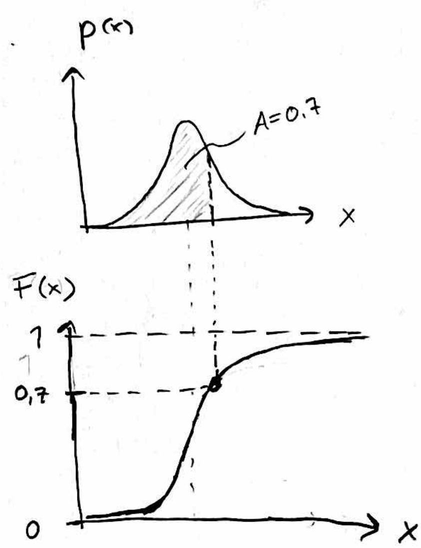
Discrete: $$\sum_{x \in D} p(x) = 1$$

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:

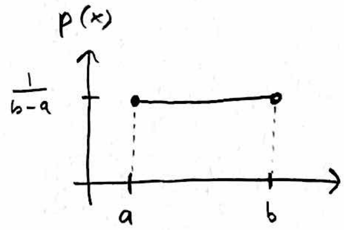
$$p(x) = \frac{d}{dx} F(x)$$

Some important (1D) prob. distributions

- The <u>uniform</u> distribution

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

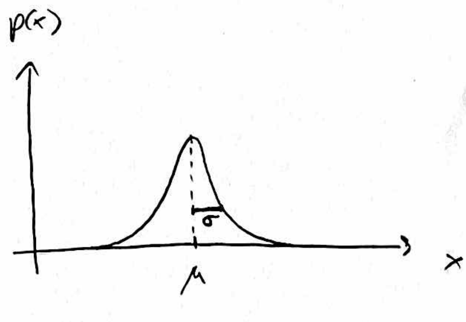
- Parameters: a, b
  $[a]=[b]=[x]$
- Standard form: $a=0$
  $b=1$



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

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

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



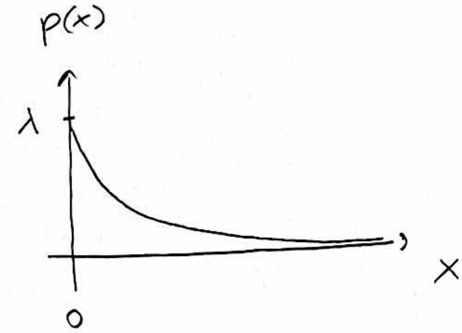
{ Pops up everywhere because of the Central Limit Theorem }

- The <u>exponential</u> distribution

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

  - Parameters : λ

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



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

A couple of discrete prob. functions

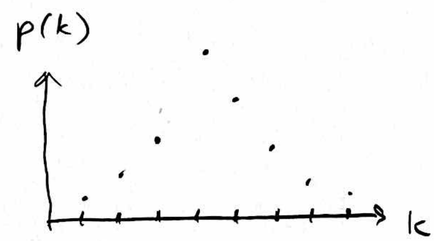
- The binomial distribution

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

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

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

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

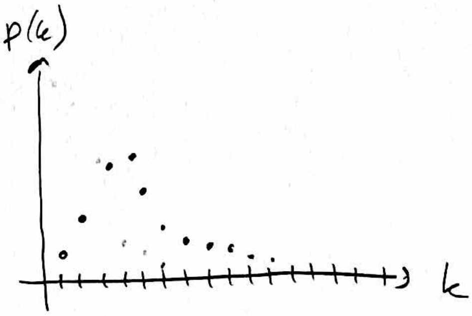


- The Poisson distr.

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

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

- 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!

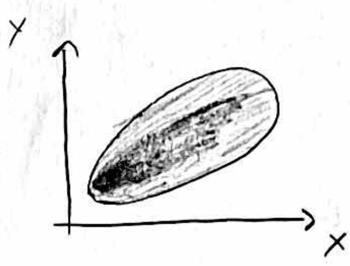
Prob. dens. functions of many variables

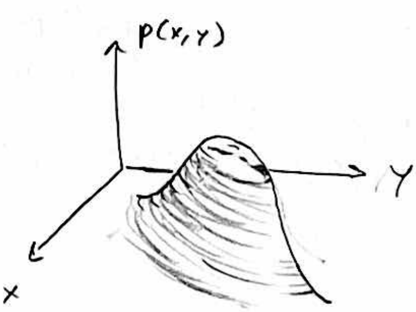
- Notation: $p(x_1, x_2, x_3, \dots)$ or $p(\bar{x})$

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





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

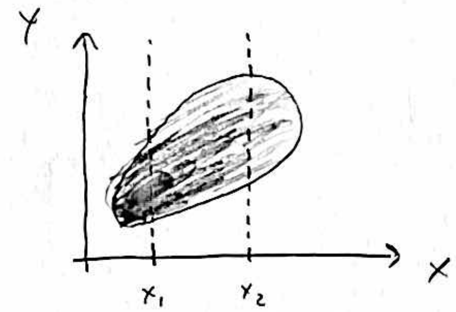
- <u>Conditional</u> pdfs

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

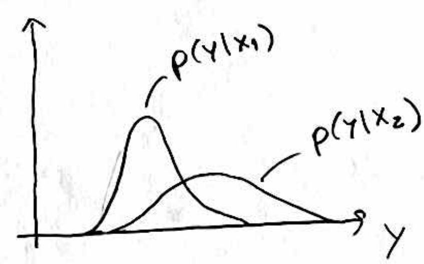
    [analogous for $p(x|y)$]

  - 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

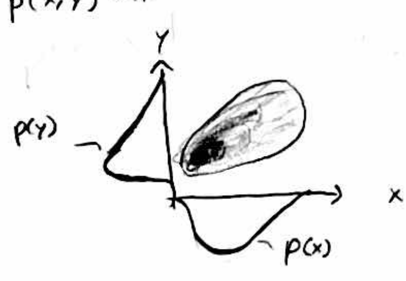


- <u>Marginal</u> pdfs

  - $p(x)\,dx$ = Prob ($X \in [x, x+dx]$, independent of Y)

  - Prob dens. over a 1D space (here: x axis) [analogous for $p(y)$]

$$p(x) = \int p(x,y)\,dy \quad \text{"marginalise over y"}$$
$$p(y) = \int p(x,y)\,dx \quad \text{" —— " —— x"}$$



- Useful relations:

Bayes theorem

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

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

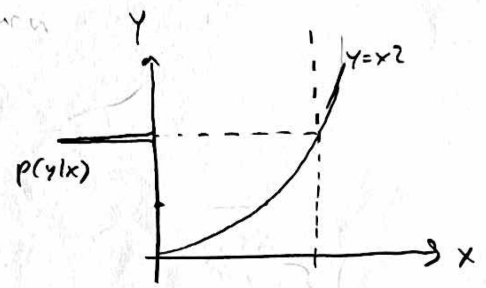
Cont.: $$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\int p(x|y)p(y)\,dy}$$

- Sometimes a "deltafunction perspective" is useful

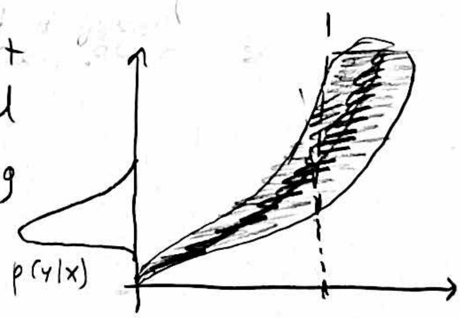
<u>Instead of</u>:
- X is rand. variable with pdf $p(x)$
- $Y = x^2$ is a function of x

<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



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

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