Modern Methods of Statistical Data Analysis

From parameter estimation to deep learning — A guided tour of probability

Lecture 3
-
Fundamental Concepts III & the Monte Carlo Method

P.-D. Dr. Roger Wolf

roger.wolf@kit.edu

Dr. Pablo Goldenzweig

pablo.goldenzweig@kit.edu

Dr. Jan Kieseler

jan.kieseler@kit.edu

Dr. Slavomira Stefkova

slavomira.stefkova@kit.edu

Karlsruher Institut für Technologie

Program today

5' break

- Recap of lecture 2
- Complete our tour of important distributions
- Independent identically distributed (i.i.d.) RVs
- Central Limit Theorem
- Answers to quiz 2

- History of Monte Carlo
- The MC method
- Generating random numbers
  - Algorithms (LCG, Mersenne Twister, …)
- Transformation method
  - Transform uniform distribution into various other distributions
- Acceptance-rejection method
- Implementation in ROOT and Python
- Quiz 3

Textbook by L. Lista

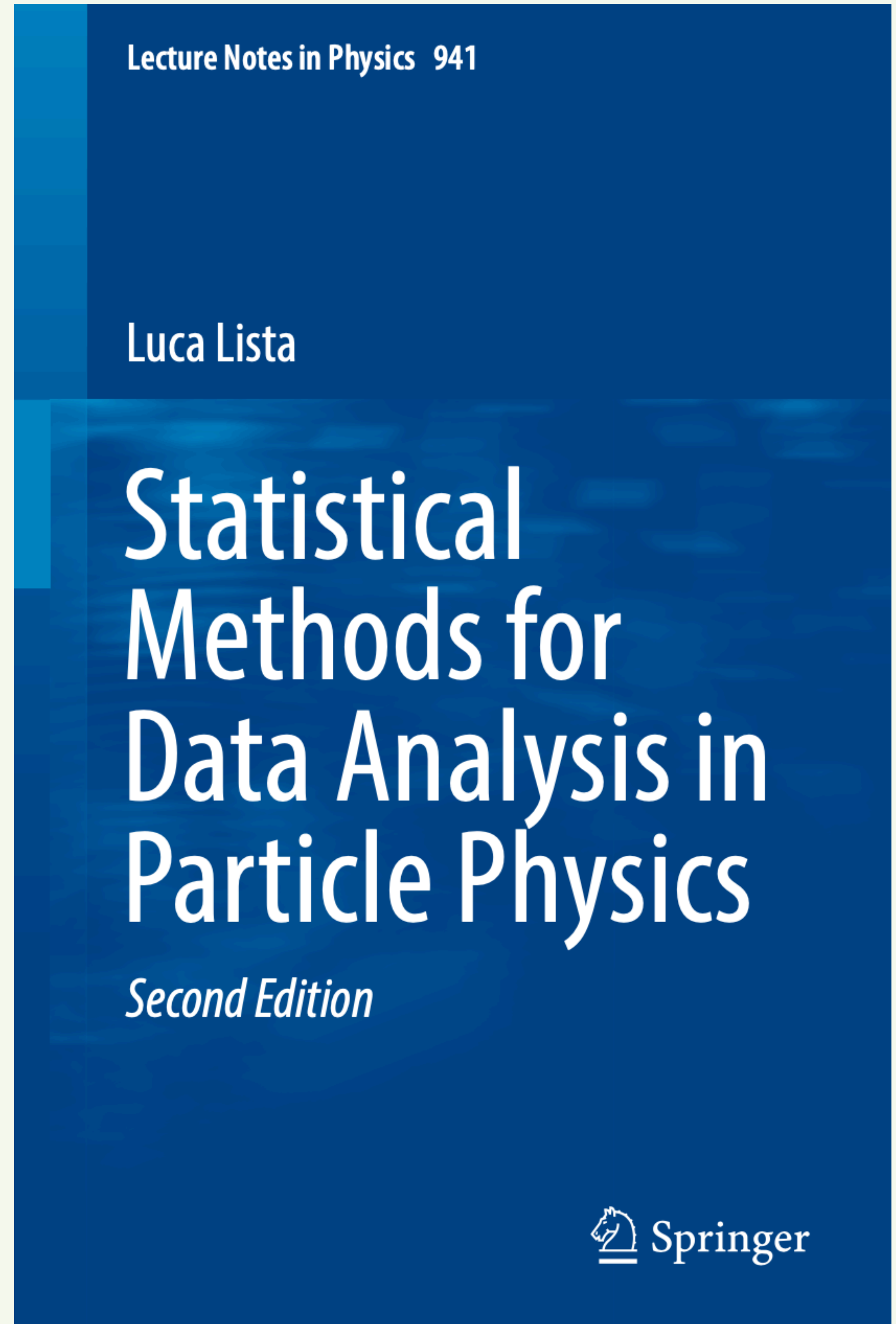
Reminder

ILIAS:

/Reading material / Textbooks / StatisticalMethodsForDataAnalysisInParticlePhysics_LLista.PDF

An excellent book which I recommend you read through (especially the Random Numbers and MC methods chapter which we'll cover today, and the later chapters on discoveries and upper limits, which we will cover soon)**.**

(See required reading slide at the end of lecture)



Independence

Review

Two events E and F are defined as independent if: $P(EF) = P(E)P(F)$

Otherwise E and F are called dependent events

If E and F are independent, then: $P(E|F) = P(E)$

Intuition through proof:

$$P(E|F) = \frac{P(EF)}{P(F)}$$

def of conditional probability (*Lecture 1, s49*)

$$= \frac{P(E)P(F)}{P(F)}$$

Independence of E and F

$$= P(E)$$

Knowing that F happened does not change our belief that E happened

Key point regarding RVs

Translate a problem statement into a **random variable**

I.e., model real life situations with ***probability distributions***

RV grid (from lecture 02)

Review

+ Uniform RV

Experiment-Fixed/known \ *Variable*	Number of successes	Time until success	
One trial	Ber (p)	Geo (p)	One success
	↑ $N = 1$	↑ $r = 1$	
Several trials	Bin (N, p)	NegBin (r, p)	Several Successes
Interval of time	Poi (ν)	Exp (τ)	Interval of time to first success

Lets add the big one…

Normal Random Variable

Normal RV

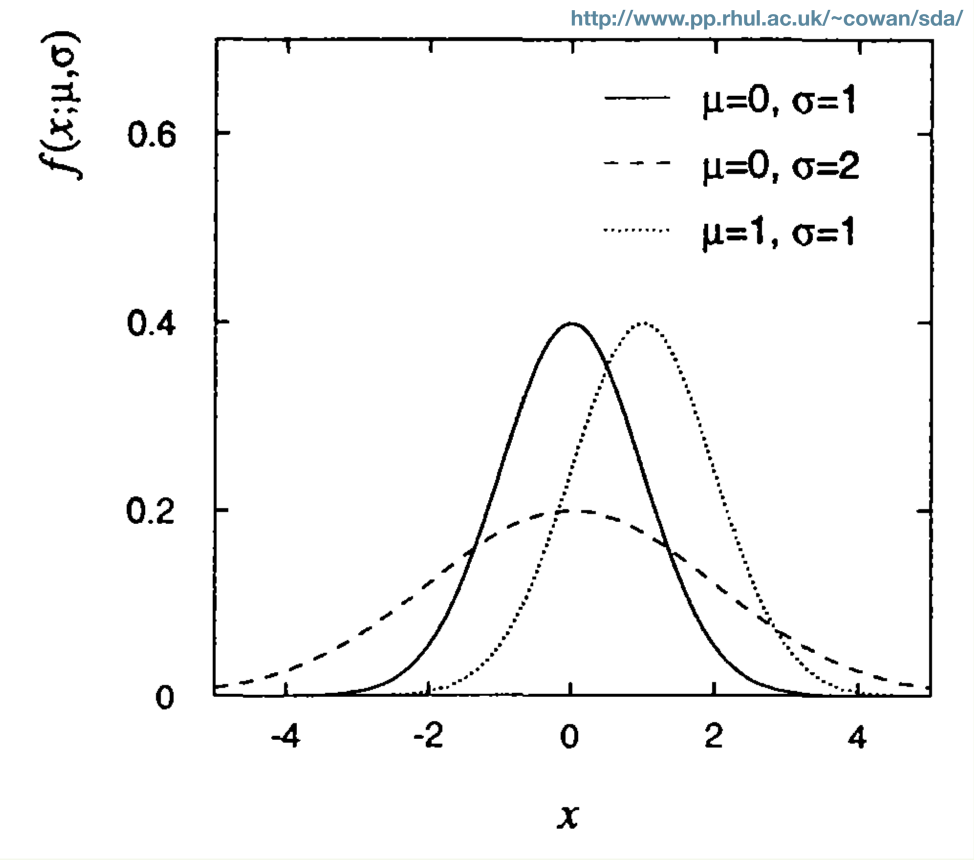
- The Gaussian (or Normal) PDF of a continuous variable x $(-\infty < x < \infty)$ is defined as

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

  - Depends on two parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$. This notation is clearly motivated by the **mean** and **variance**:

$$E[X] = \int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)dx = \boldsymbol{\mu}$$

$$V[X] = \int_{-\infty}^{\infty} (x-\mu)^2\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)dx = \boldsymbol{\sigma^2}$$



Why the Normal?

also..

Because it's easy to use!

(but please stay critical of how to model real-world phenomena)

- Common for natural phenomena:
  - Height, weight, etc.

Actually, log-normal

- Most noise in the world is Normal

Just an assumption

- Often results from the sum of many random variables

Only if equally weighted

- Sample means are distributed normally

- *A Normal maximizes entropy (i.e., randomness) for a given mean and variance*

ILIAS: /Reading material / L03 /WhyTheNormalDistribution?, InfoTheoryAndMaxEntropy

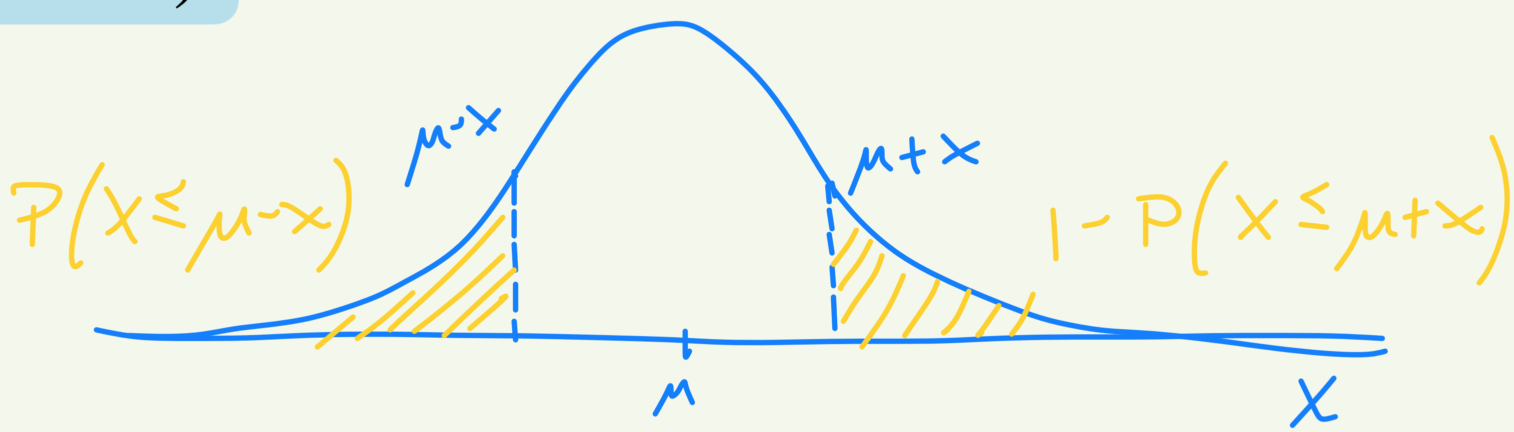
The **normal distribution** seems to be the center of the galaxy of distributions towards which all other distributions gravitate

E. T. Jaynes, *Probability Theory: the Logic of Science*, Cambridge University Press, 2003

Properties of Normal RVs

- Linear transformations of Normal RVs are also Normal RVs
  - If $Y = aX + b$, then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$
    - Proof:
      - $E[Y] =$ $E[aX + b] = aE[X] + b = a\mu + b$ **Linearity of expectation**
      - $\mathrm{Var}[Y] =$ $\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X] = a^2\sigma^2$ **Variance is not linear**
      - Y is also Normal [Ross, section 2.3.4, page 34]

- The PDF of a Normal RV is symmetric about the mean μ
  - $F(\mu - x) = 1 - F(\mu + x)$



Piece by piece

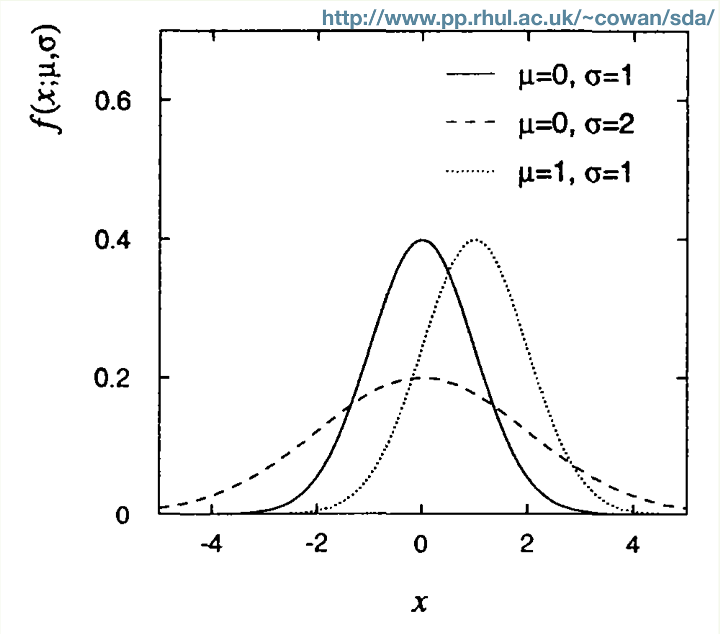
$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Symmetric around μ

Variance σ^2 manages spread

Exponential tail

Normalization constant



A fun read for history buffs:

ILIAS: /Reading material / L03 / TheEvolutionOfTheNormalDistribution

Let's use it

- You spend some minutes, X, cycling between classes
  - Average time spent: $\mu = 4$ minutes
  - Variance of time spent: $\sigma^2 = 2$ minutes^2
- Suppose X is normally distributed. What is the probability you spend ≥ 6 minutes traveling?

$$X \sim \mathcal{N}(\mu = 4, \sigma^2 = 2)$$

$$P(X \geq 6) = \int_6^\infty f(x)dx$$

$$= \int_6^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad = \ldots?$$

Don't try too hard…

Cannot be solved analytically

Computing probabilities with Normal RVs

- For a normal RV $X \sim \mathcal{N}(\mu, \sigma^2)$, its CDF has no closed form

$$P(X \le x) = F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

- However, can solve for probabilities numerically using a function Φ

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

CDF of the Standard (unit) Normal, Z

Expectation: $E[Z] = \mu = 0$

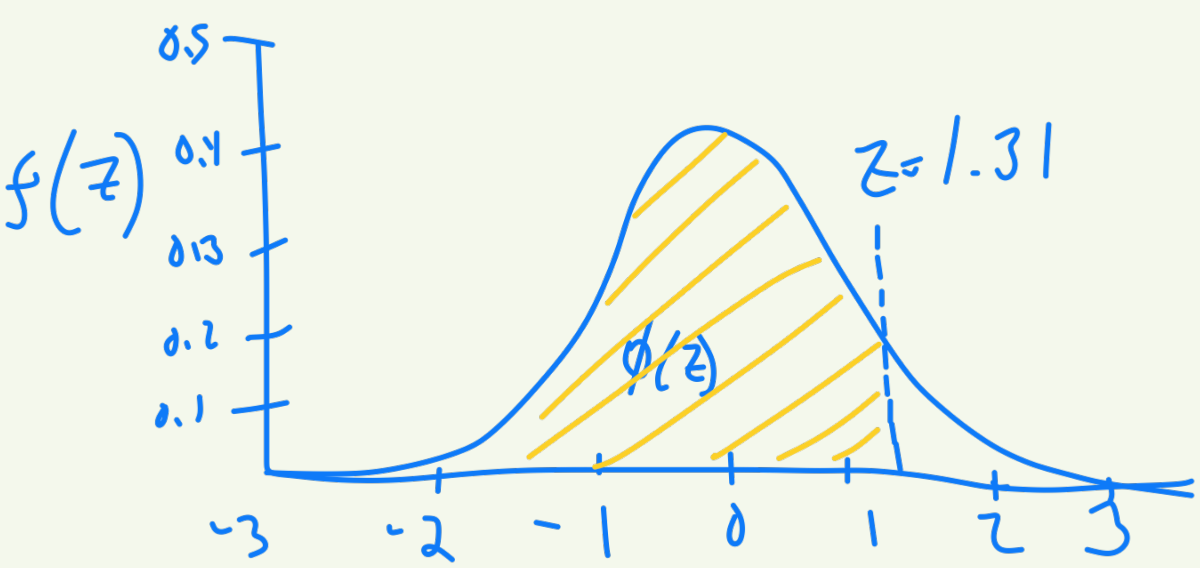
Variance: $\mathrm{Var}[Z] = \sigma^2 = 1$

CDF of Z

- Defined as: $P(Z \leq z) = \Phi(z)$
  - Φ has been numerically computed

e.g.,

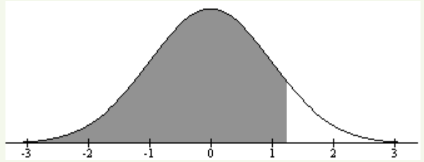
$$P(Z \leq 1.31) = \Phi(1.31)$$



Standard Normal Table only has probabilities $\Phi(z)$ for $z \geq 0$

Standard Normal Table

Note: An entry in the table is the area under the curve to the left of z, $P(Z \leq z) = \Phi(z)$



z	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	0.6700	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	0.7703	0.7734	0.7764	0.7793	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	0.8906	0.8925	0.8943	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	0.9995	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

First Standard Normal Table

TABLE PREMIÈRE.

Intégrales de $e^{-tt}\,dt$, depuis une valeur quelconque de t jusqu'à t infinie.

t	Intégrale.	Diff. prem.	Diff. II.	Diff. III.
0,00	0,88622692	999968	201	199
0,01	0,87622724	999767	400	199
0,02	0,86622957	999367	599	200
0,03	0,85623590	998768	799	199
0,04	0,84624822	997969	998	197
0,05	0,83626853	996971	1195	199
0,06	0,82629882	995776	1394	196

Computed by Christian Kramp, French astronomer (1760-1826), in *Analyse des Refractions Astronomiques et Terrestres*, 1799

Used a Taylor series expansion to the third power

$$\int_{0.03}^{\infty} e^{-x^2} dx = 0.856236$$

Let's try again

- You spend some minutes, X, cycling between classes
  - Average time spent: $\mu = 4$ minutes
  - Variance of time spent: $\sigma^2 = 2$ minutes^2

- Suppose X is normally distributed. What is the probability you spend ≥ 6 minutes traveling?

$$X \sim \mathcal{N}(\mu = 4, \sigma^2 = 2)$$

✗ $P(X \geq 6) = \int_6^\infty f(x)dx$ (no analytic solution)

1. Compute $z = \frac{(x-\mu)}{\sigma}$

$$P(X \geq 6) = 1 - F(6)$$
$$= 1 - \Phi\left(\frac{6-4}{\sqrt{2}}\right)$$
$$\approx 1 - \Phi(1.41)$$

2. Look up $\Phi(z)$ in table

z	0.00	0.01
0.0	0.5000	0.5040
0.1	0.5398	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217
0.4	0.6554	0.6591
0.5	0.6915	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611
0.8	0.7881	0.7910
0.9	0.8159	0.8186
1.0	0.8413	0.8438
1.1	0.8643	0.8665
1.2	0.8849	0.8869
1.3	0.9032	0.9049
1.4	0.9192	0.9207

$$1 - \Phi(1.41)$$
$$\approx 1 - 0.9207$$
$$= 0.0793$$

Is there an easier way?

Use Python

```
from scipy import stats
X = stats.norm(mu, std)
X.cdf(x)
```

Who gets to approximate?

# of successes in N independent trials w.p. of success p

$$n \sim \mathrm{Bin}(N, p)$$

$$E[n] = Np$$

$$\mathrm{Var}[n] = Np(1-p)$$

- Computing probabilities on Binomial RVs can be computationally expensive
- Two reasonable approximations, ***but when to use which?***

?

$$Y \sim \mathrm{Poi}(\nu)$$

$$\nu = Np$$

N large (> 20)

p small (< 0.05)

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

$$\mu = Np$$

$$\sigma^2 = Np(1-p)$$

N large (> 20)

p mid-range ($Np(1-p) > 10$)

i.i.d. RVs & the CLT

Independence of multiple RVs

Review

n events $E_1, E_2, \ldots, E_n$ are independent if:

For $r = 1,\ldots,n$:

for every subset $E_1, E_2, \ldots, E_r$:

$$P(E_1, E_2, \ldots, E_r) = P(E_1)P(E_2)\cdots P(E_r)$$

We have **independence** of n discrete RVs $X_1, X_2, \ldots, X_n$ if for all $x_1, x_2, \ldots, x_n$:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_i^n P(X_i = x_i)$$

Shorthand notation →

$$P_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_i^n P_{X_i}(x_i)$$

Shorter... →

$$P(x_1, x_2, \ldots, x_n) = \prod_i^n P(x_i)$$

We have **independence** of n continuous RVs $X_1, X_2, \ldots, X_n$ if for all $x_1, x_2, \ldots, x_n$:

$$P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) = \prod_i^n P(X_i \le x_i)$$

Shorthand notation →

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_i^n f_{X_i}(x_i)$$

Shorter... →

$$f(x_1, x_2, \ldots, x_n) = \prod_i^n f(x_i)$$

i.i.d. random variables

- Consider n variables $X_1, X_2, \ldots, X_n$
  - $X_1, X_2, \ldots, X_n$ are **independent and identically distributed (i.i.d.)** if
    - $X_1, X_2, \ldots, X_n$ are **independent**, and
    - all have the **same PMF** *(if discrete)* or **PDF** *(if continuous)*
      - $E[X_i] = \mu$ for $i = 1, \ldots, n$
      - $\mathrm{Var}[X_i] = \sigma^2$ for $i = 1, \ldots, n$

Quick check: Are $X_1, X_2, \ldots, X_n$ i.i.d. with the following distributions?

1. $X_i \sim \mathrm{Exp}(\tau)$, X_i independent ✓
2. $X_i \sim \mathrm{Exp}(\tau_i)$, X_i independent ✗ (unless τ_i equal)
3. $X_i \sim \mathrm{Exp}(\tau)$, $X_1 = X_2 = \cdots = X_n$ ✗ dependent! ($x_1 = x_2 = \ldots = x_n$)
4. $X_i \sim \mathrm{Bin}(n_i, p)$, X_i independent ✗ (unless n_i equal)

In physics terms

1. Assume physics and experimental conditions stay the same.

2. Assume each event (collision/measurement) has no influence over others.

As always in physics:

Pretend this is true and let uncertainties take care of the rest

Central Limit Theorem

Arguably the most important result in statistics

Consider $\boldsymbol{n}$ independent and identically distributed (i.i.d.) variables $\boldsymbol{X_1, X_2, \ldots, X_n}$ with $\boldsymbol{E[X_i] = \mu}$ and $\mathbf{Var}[\boldsymbol{X_i}] = \boldsymbol{\sigma^2}$

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

As $n \to \infty$

I.e.,

- The sum of $\boldsymbol{n}$ independent continuous RVs $\boldsymbol{X_i}$ with means $\boldsymbol{\mu_i}$ and variances $\boldsymbol{\sigma_i^2}$ become a Gaussian RV with mean and variance

$$\mu = \sum_{i=1}^{n} \mu_i \qquad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2$$

As $n \to \infty$

- This holds under fairly general conditions **regardless of the form of the individual PDFs of the other** X_i

Simple ex.: *Sum of dice rolls*

Roll n independent dice

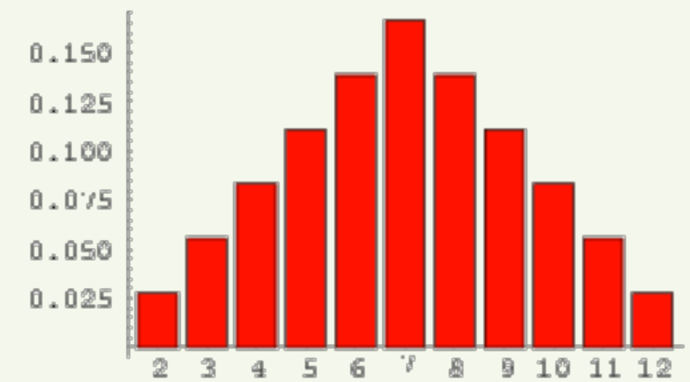
Let X_i be the outcome of roll i

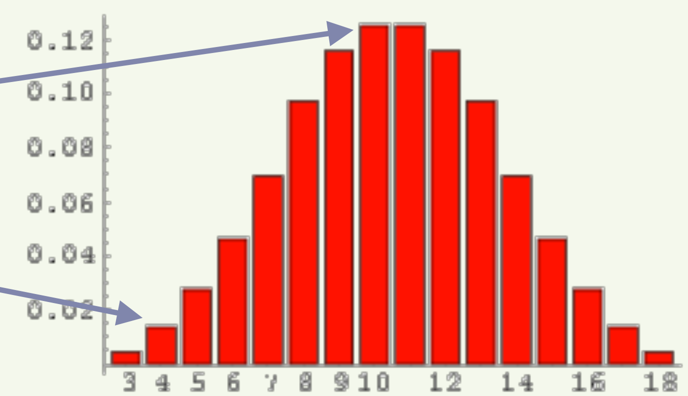
X_i are i.i.d.

How many ways can you roll a total of 4 vs. 10?

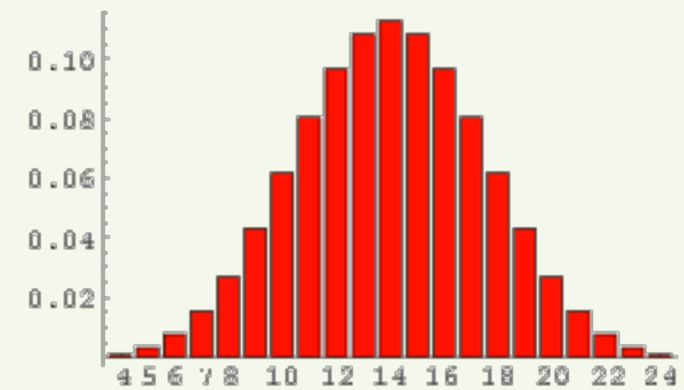
$\sum_{i=1}^{1} X_i$ **Sum of 1 die roll**

$\sum_{i=1}^{2} X_i$ **Sum of 2 dice rolls**





$\sum_{i=1}^{3} X_i$ **Sum of 3 dice rolls**



$\sum_{i=1}^{4} X_i$ **Sum of 4 dice rolls**

As $n \to \infty$

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

What about the average of i.i.d. RVs?

(i.e., sample mean)

Let $X_1, X_2, \ldots, X_n$ be i.i.d., where $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2$. *As* $n \to \infty$*:*

Define: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **(sample mean)** $Y = \sum_{i=1}^{n} X_i$ **(sum)**

$$Y \sim \mathcal{N}(n\mu, n\sigma^2)$$ **(CLT as $n \to \infty$)**

$$\bar{X} = \frac{1}{n}Y$$

$$\bar{X} \sim \mathcal{N}(?, ?)$$

$$E[\bar{X}] = \frac{1}{n}E[Y] = \frac{1}{n} \cdot n\mu = \mu$$

$$\text{Var}[\bar{X}] = \left(\frac{1}{n}\right)^2 \text{Var}[Y] = \left(\frac{1}{n}\right)^2 n\sigma^2 = \frac{\sigma^2}{n}$$

The average of i.i.d. RVs is normally distributed with mean μ and variance σ^2/n

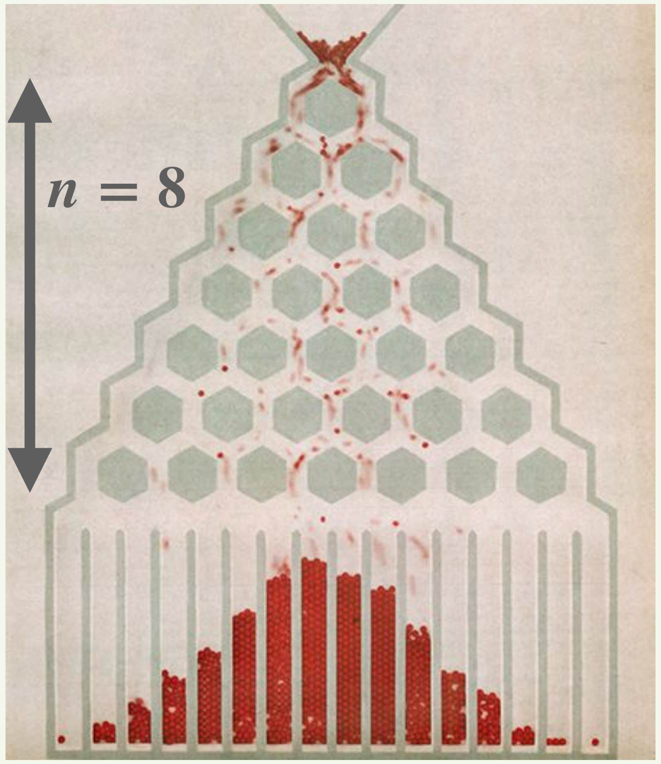
"I know of scarcely anything so apt to impress the imagination as the wonderful **form of cosmic order** expressed by the **Central limit theorem**. The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshaled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along."

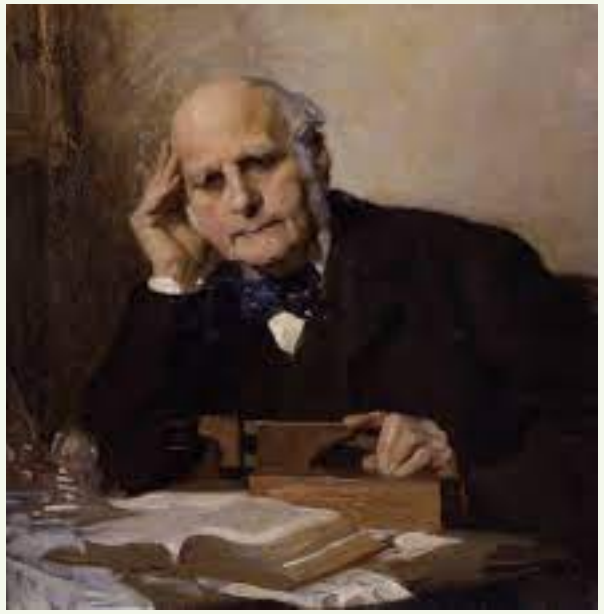
—Sir Francis Galton

(Yes, the Galton Board)

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

As $n \to \infty$





Putting it all together

Let $X_1, X_2, \ldots, X_n$ be i.i.d., where $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2$. *As* $n \to \infty$*:*

Sum of i.i.d. RVs

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

Working with the CLT

Average of i.i.d. RVs (sample mean)

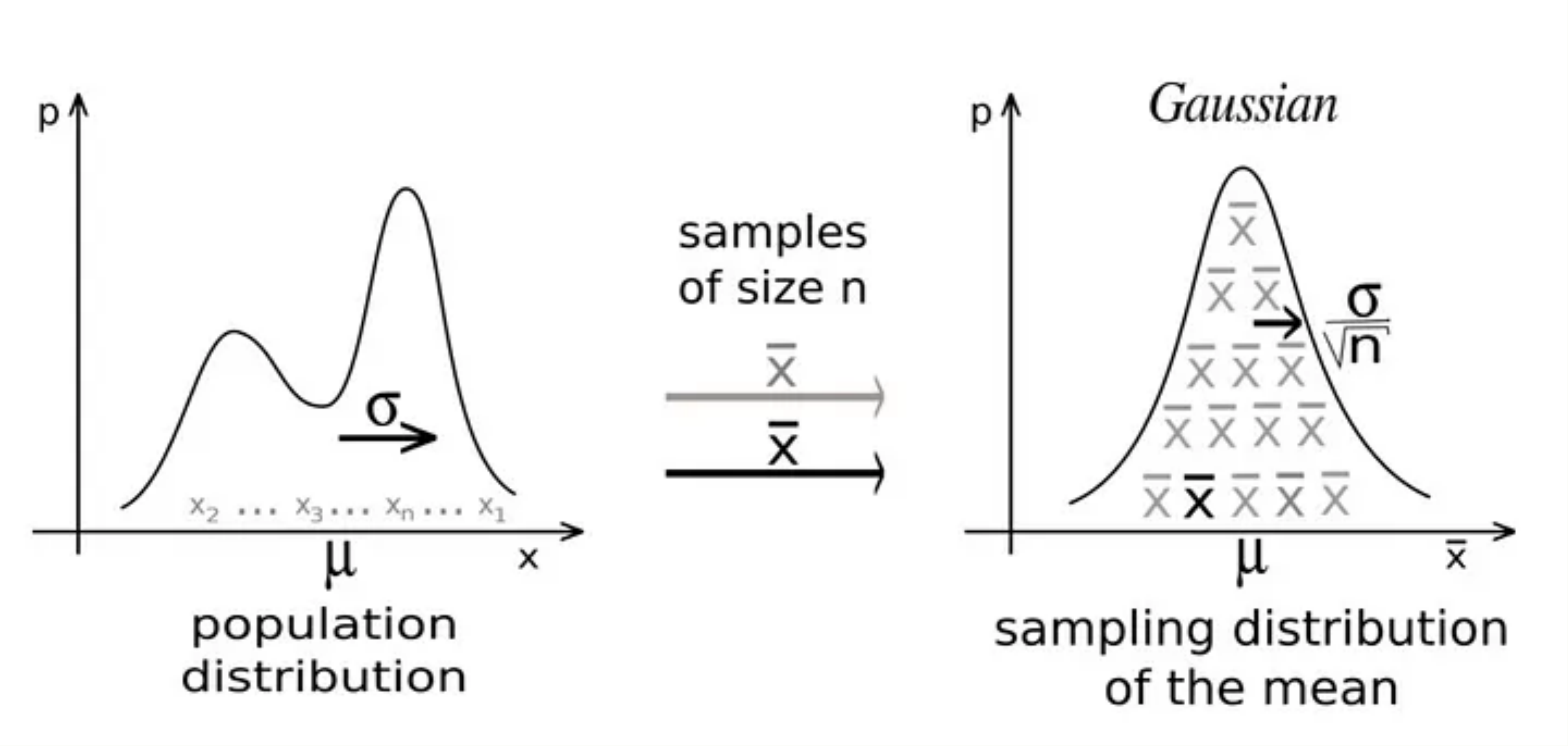
$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Interpret: As we increase n (the size of our sample):

- The variance of our sample mean σ^2/n decreases
- The probability that our sample mean $\bar{X}$ is close to the true mean μ increases

Key take home message

No matter what the distribution of the population is, the distribution of mean samples from the population will always be Normally distributed



i.e., No matter what the distribution of the sample is, if you sample batches of data from that distribution and take the mean of each batch, *the mean values from those batches will be Normally distributed*

Proof of the CLT

- The Fourier Transform of a PDF is called a characteristic function
- Take the characteristic function of the probability mass of the sample distance from the mean, divided by the SD $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma}$
- Show that this approaches an exponential function in the limit as $n \to \infty$ $f(x) = e^{-x^2/2}$
- This function is in turn the characteristic function of the Standard Normal, $Z \sim \mathcal{N}(0,1)$

Proof: Cowan, 10.1-10.3 (pages 143-149)
Ross, 2.8 (pages 82-83)

Question from last lecture:

Do errors need to be Gaussian for propagation of errors to hold?

- In general, error propagation is founded on the assumption that:
  - The error is small (where the scale for smallness is set by the ratio of 1st to 2nd derivatives) compared to the value of the quantity (otherwise we can't use the Taylor expansion);
  - The measurement errors in the input variables are independent, & the measurement errors are independent from one measurement to the next;
  - There are many measurements of each variable.
- If you have a sufficient # of variables with small but non-Gaussian errors, the CLT says that the result will be Gaussian distributed (here you can compute the std. dev. of the non-Gaussian individual distributions, and use them as a surrogate in your error propagation calculation).
- If, however, you have a small # of variables AND the distribution is non-Gaussian, you need to perform a MC simulation:
  - Sample the distribution of your input variables, transform them according to the error propagation formula, plot the resulting output distribution, and compute it's shape.

Next time

Preview

- **Central Limit Theorem:** ✓
  - Sample mean $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$
  - If we know μ and σ^2, we can compute probabilities on the sample mean $\bar{X}$ of a given sample size n

- **In real life:**
  - Yes, the CLT still holds…
  - But we often don't know μ or σ^2 of our original distribution
  - However, we can collect data (a sample of size n)
  - **Question: How can we estimate the values μ or σ^2 from our sample?**
    - *Answer: Covered in next lecture on parameter estimation*

Right now let's take a tour of the Monte Carlo method

We'll need this as well

A few more continuous RVs for you to review at home…

Covered in Cowan sections 2.6-2.9

Multi-dimensional Gaussian

- The N-dimensional generalization of the Gaussian distribution is defined by

$$f(\boldsymbol{x};\boldsymbol{\mu},V) = \frac{1}{(2\pi)^{N/2}|V|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

  - Here $\boldsymbol{x}$ and $\boldsymbol{\mu}$ are column vectors containing $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ and $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_N$.
  - $|V|$ is the determinant of the symmetric $N \times N$ covariance matrix V
  - Expectation values and (co)variances are:

$$\begin{aligned} E[x_i] &= \mu_i \\ V[x_i] &= V_{ii} \\ \mathrm{cov}[x_i, x_j] &= V_{ij}. \end{aligned}$$

- In 2D:

$$f(x_1, x_2; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\}$$

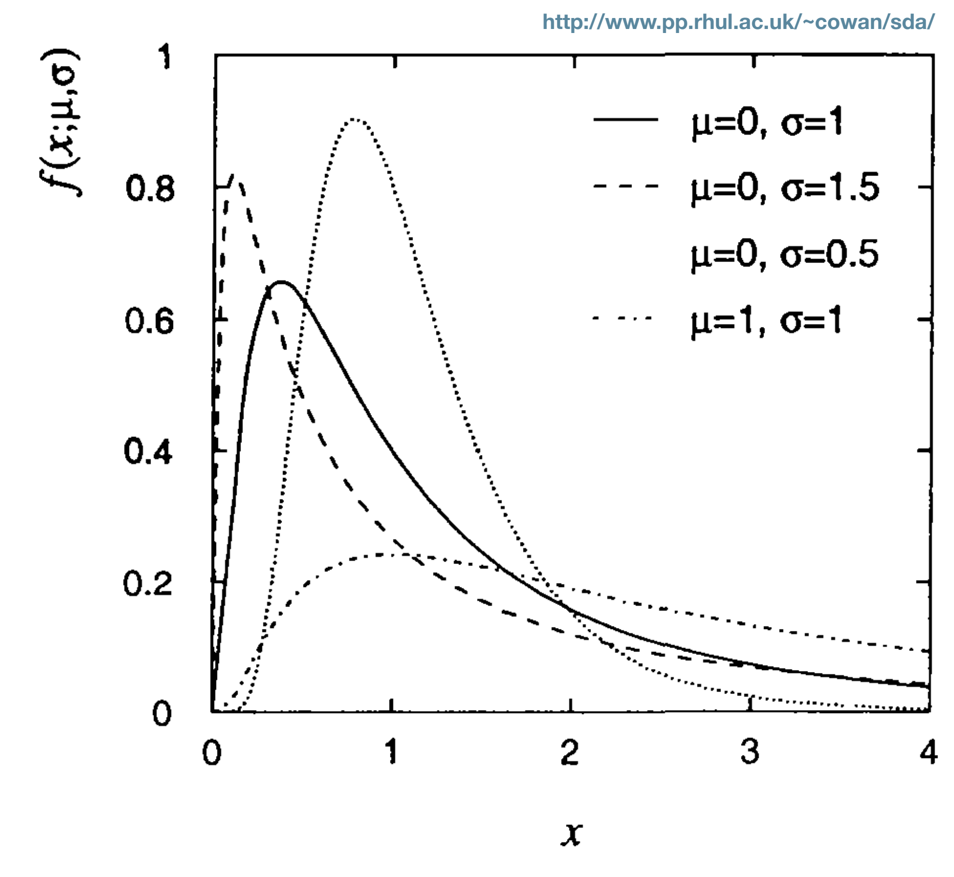
Correlation coefficient: $\rho = \mathrm{cov}[x_1, x_2]/(\sigma_1\sigma_2)$

Log-normal distribution

- If a continuous **variable *y*** is Gaussian with mean μ and variance σ^2, then **x = exp(y)** follows the **log-normal** distribution:

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

- Expectation value and variance: $E[x] = \exp(\mu + \frac{1}{2}\sigma^2)$, $V[x] = \exp(2\mu+\sigma^2)[\exp(\sigma^2)-1]$.
- **Note:** in this notation μ and σ^2 are not the mean and variance of x, but of the corresponding Gaussian distribution for log x.



Chi-square distribution (i)

- The χ^2 (chi-square) distribution of the continuous variable z $(0 < z < \infty)$ is defined by

$$f(z;n) = \frac{1}{2^{n/2}\,\Gamma(n/2)} z^{n/2-1} e^{-z/2}, \qquad n = 1, 2, \ldots,$$

- The parameter n is called **number of degrees of freedom** and the gamma function:

$$\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, dt$$

To calculate χ^2, need to know:
$\Gamma(n) = (n-1)!$ for integer n,
$\Gamma(x+1) = x\Gamma(x)$ and $\Gamma(1/2) = \sqrt{\pi}$

- Expectation value and variance:

$$E[z] = \int_0^\infty z \frac{1}{2^{n/2}\,\Gamma(n/2)} z^{n/2-1} e^{-z/2}\, dz = n$$

$$V[z] = \int_0^\infty (z-n)^2 \frac{1}{2^{n/2}\,\Gamma(n/2)} z^{n/2-1} e^{-z/2}\, dz = 2n$$

Chi-square distribution (ii)

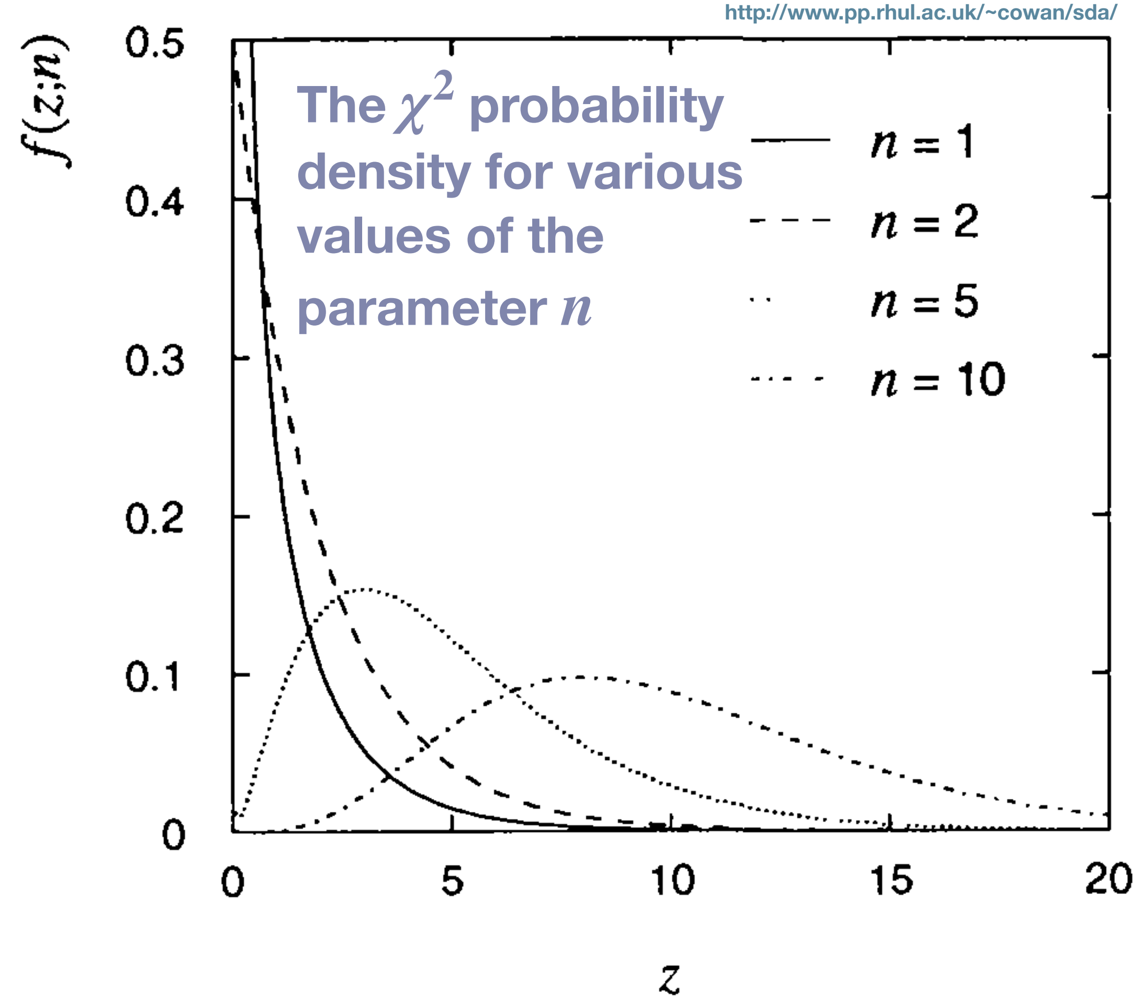
- The χ^2 **distribution** is important due to its **relation** to the **sum of squares of Gaussian distributed random variables**. Given N independent Gaussian random variables x_i with known means $\boldsymbol{\mu_i}$ and variances $\boldsymbol{\sigma_i^2}$, the variable

$$z = \sum_{i=1}^{N} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

is distributed like a χ^2 **distribution with N degrees of freedom.**

Proof in Cowan Sec. 10.2

http://www.pp.rhul.ac.uk/~cowan/sda/



Also holds if x_i are **not independent** but are N-dimensionally Gaussian distributed

$$z = (\boldsymbol{x} - \boldsymbol{\mu})^T V^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$$

*Variables following a χ^2 **distribution will play an important role in tests of goodness-of-fits!***

Cauchy (Breit-Wigner) distribution

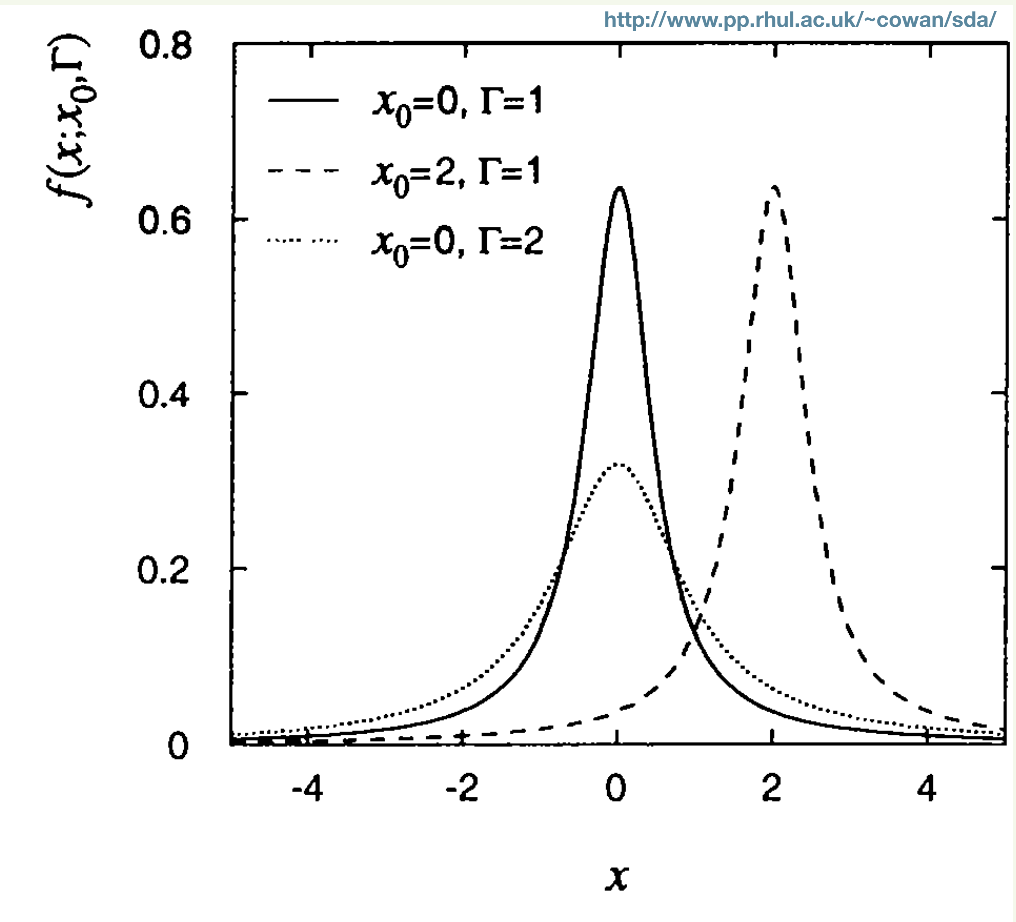
- Cauchy or Breit-Wigner PDF of a continuous variable x $(\infty < x < \infty)$ is defined by

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

- Special case of Breit-Wigner distribution encountered in particle physics

$$f(x;\Gamma,x_0) = \frac{1}{\pi}\frac{\Gamma/2}{\Gamma^2/4+(x-x_0)^2},$$

particle mass: x_0
particle width: Γ



- The expectation value and variance are not well defined for this distribution as the integrals

$$\int_{-\infty}^{0} x\, f(x)\mathrm{d}x \quad \text{and} \quad \int_{0}^{\infty} x\, f(x)\mathrm{d}x$$ are divergent

- Use x_0 (= Mode) and Γ (= FWHM) to give information about the PDF.

Random variables in Python

- Paket numpy.random:
  - binomial(n, p[, size])
  - chisquare(df[, size])
  - exponential([scale, size])
  - lognormal([mean,
  - sigma, size])
  - multinomial(n, pvals[, size])
  - multivariate_normal(mean, cov[, size])
  - normal([loc, scale, size])
  - poisson([lam, size])
  - power(a[, size])
  - standard_cauchy([size])
  - standard_exponential([size])
  - standard_normal([size])
  - standard_t(df[, size])
  - triangular(left, mode, right[, size])
  - uniform([low, high, size])

https://docs.scipy.org/doc/numpy/reference/routines.random.html

Answer Time: Quiz 2

1. Error propagation: You have 2 random variables x_1, x_2 with a known covariance matrix

$$C = \begin{pmatrix} 0.5. & 0.2 \\ 0.2 & 0.7 \end{pmatrix},$$

and expectation values $\mu_1 = 6$ and $\mu_2 = 1$. You know want to determine the covariance of two functions f_1 and f_2 of x_1 and x_2 defined as

$$f_1(x_1, x_2) = x_1 + x_2\,,$$
$$f_2(x_1, x_2) = \sqrt{x_1^2 + x_2^2}\,.$$

Calculate the Jacobian $A_{ij} = \left[\frac{\partial f_i}{\partial x_j}\right]_{x_1=\mu_1, x_2=\mu_2}$ and the covariance matrix D between f_1 and f_2. What is the **correlation** between f_1 and f_2? *Hint:* Use $D = ACA^T$.

- Jacobian:

$$A = \begin{pmatrix} 1 & 1 \\ \frac{x_1}{\sqrt{x_1^2+x_2^2}} & \frac{x_2}{\sqrt{x_1^2+x_2^2}} \end{pmatrix}\Bigg|_{x_1=\mu_1, x_2=\mu_2}$$

- New covariance:

$$D = \begin{pmatrix} 1 & 1 \\ \frac{\mu_1}{\sqrt{\mu_1^2+\mu_2^2}} & \frac{\mu_2}{\sqrt{\mu_1^2+\mu_2^2}} \end{pmatrix} \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.7 \end{pmatrix} \begin{pmatrix} 1 & \frac{\mu_1}{\sqrt{\mu_1^2+\mu_2^2}} \\ 1 & \frac{\mu_2}{\sqrt{\mu_1^2+\mu_2^2}} \end{pmatrix}$$

$$= \begin{pmatrix} 1.6 & 0.84 \\ 0.84 & 0.57 \end{pmatrix}$$

- Correlation:

$$\rho_{12} = D_{12}/\sqrt{D_{11}}/\sqrt{D_{22}} = 0.877\ldots$$

2. Show that the expectation value and the variance of a Poisson random variable is given by the Poisson parameter ν, i.e. that

$$E[n] = \sum_{n=0}^{\infty} n \frac{\nu^n}{n!} e^{-\nu} \doteq \nu \,,$$

$$V[n] = \sum_{n=0}^{\infty} (n-\nu)^2 \frac{\nu^n}{n!} e^{-\nu} \doteq \nu \,.$$

Hint: Use $V[n] = E[n^2] - (E[n])^2$ and $E[n^2] = E[n\,(n-1)+n] = E[n\,(n-1)] + E[n]$ and the properties of the exponential series ($e^{-x} = \sum_{n=0}^{\infty} \frac{x^n}{n!}$).

Property 1 of variance (L02, S35)

Recall we said it is often easier to compute than the definition

Expectation of a sum = sum of expectation (L02, S31)

Important property of expectation

Solution: (sorry about the difference in notation) $\nu \to \lambda,\ n \to X,\ E[n] = E(X)$

Expected value and variance of Poisson random variables. We said that λ is the expected value of a Poisson(λ) random variable, but did not prove it. We did not (yet) say what the variance was. For the expected value, we calculate, for X that is a Poisson(λ) random variable:

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} && \text{since the } x=0 \text{ term is itself 0} \\
&= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!} && \text{divided on top and bottom by } x \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} && \text{factor out } e^{-\lambda} \text{ and } \lambda \text{ too} \\
&= \lambda e^{-\lambda} \left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \dots \right) \\
&= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}$$

So in summary $E(X) = \lambda$. For $\mathrm{Var}(X) = E(X^2) - (E(X))^2 = E((X)(X-1)+X) - (E(X))^2 = E((X)(X-1)) + E(X) - (E(X))^2 = E((X)(X-1)) + \lambda - \lambda^2$. Now we calculate

$$\begin{aligned}
E((X)(X-1)) &= \sum_{x=0}^{\infty} (x)(x-1) \frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=2}^{\infty} (x)(x-1) \frac{e^{-\lambda}\lambda^x}{x!} && \text{because } x=0 \text{ and } x=1 \text{ terms are themselves 0} \\
&= \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-2)!} && \text{divide out by } x \text{ and } x-1 \\
&= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} && \text{factor out } e^{-\lambda} \text{ and } \lambda^2 \\
&= \lambda^2 e^{-\lambda} \left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \dots \right) \\
&= \lambda^2 e^{-\lambda} e^{\lambda} \\
&= \lambda^2
\end{aligned}$$

In summary, $\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

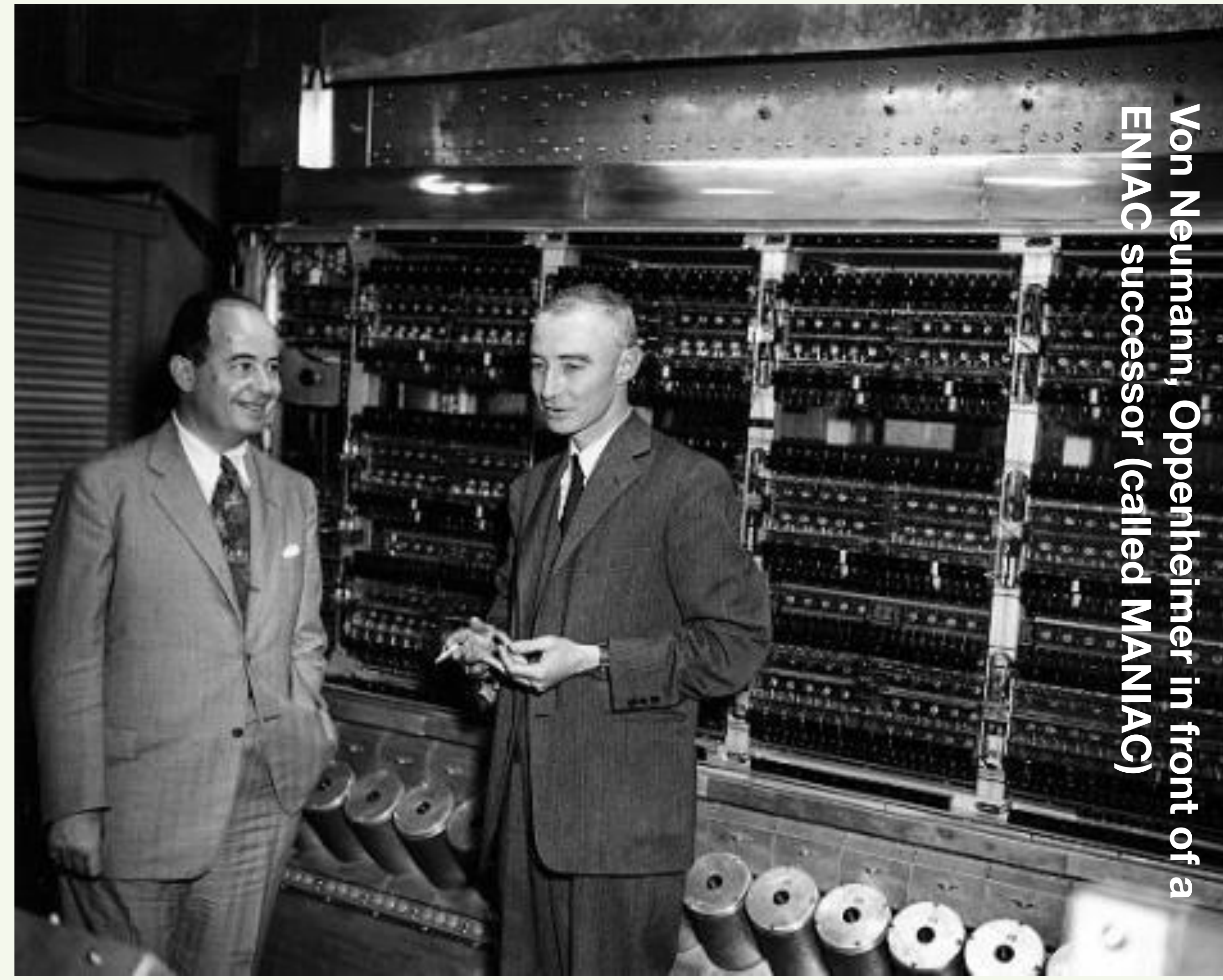
So both the expected value and the variance of X are equal to λ.

Take 5

A brief history of Monte Carlo

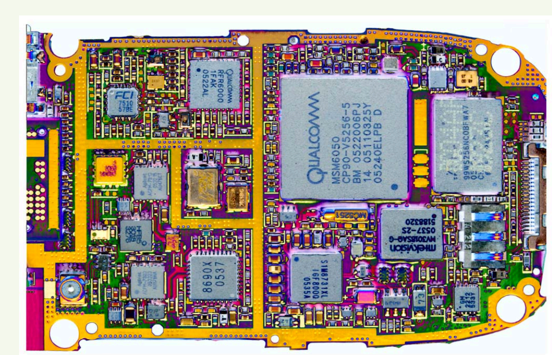
A brief history of Monte Carlo

- In 1945 two earthshaking events took place:
  - The first nuclear bomb was detonated in the Alamogordo dessert
  - The first electronic computer was built
    - **ENIAC** (= Electronic Numerical Integrator and Computer)
      - 20k vacuum tubes, 7200 crystal diodes, 5M hand-soldered connections
      - Total weight: 27 tons on 167 m^2
    - *Your cell phone: is easily >100k times more powerful whilst consuming 400k times less power*

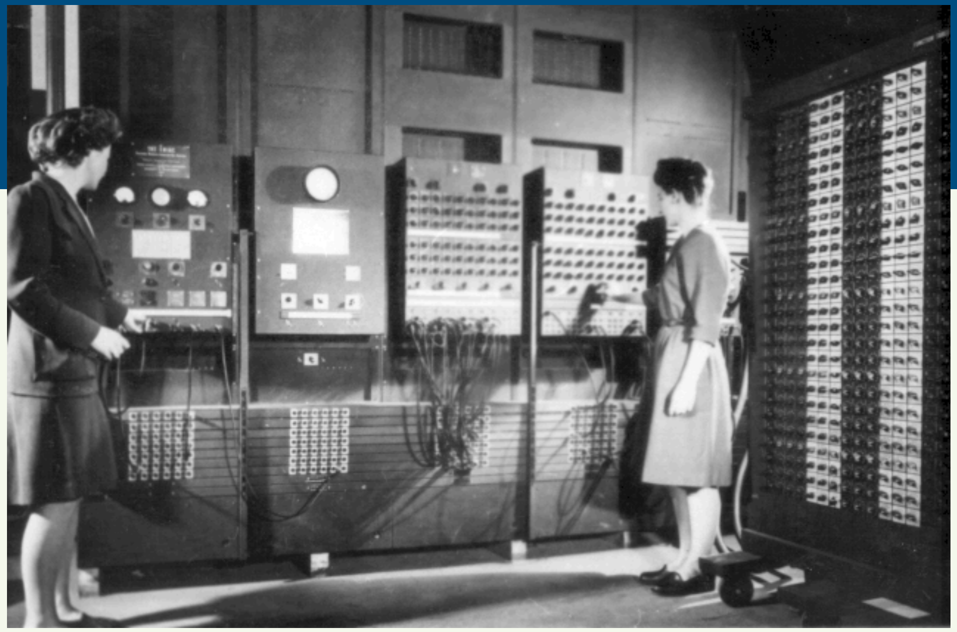


Von Neumann, Oppenheimer in front of a ENIAC successor (called MANIAC)

385 multiplication operations **per second**; five of the accumulators were controlled by a special divider/square-rooter unit to perform up to **40 division** operations **per second** or **three square root** operations **per second**.



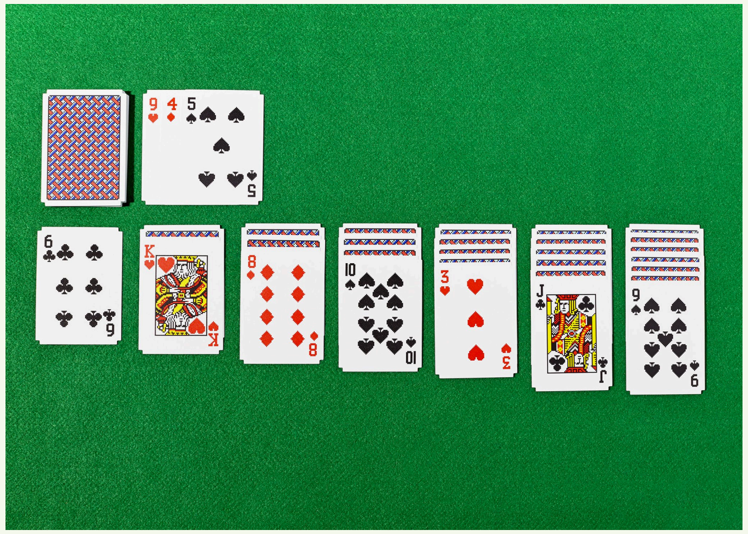
ENIAC



- First programmers:
  - Kay McNulty, Betty Jennings, Betty Snyder, Marlyn Meltzer, Fran Bilas, and Ruth Lichterman
- First test problem: related to the the hydrogen bomb
  - ENIAC's role in this made **Monte Carlo (MC)** methods a popular tool to solve physics problems
    - At the time scientists used massive groups to carry out calculations ('computers') to investigate the distance neutrons would likely travel through various materials (i.e., through an exploding atomic bomb)
      - Easy to solve numerically, very hard analytically
    - John von Neumann and Stansilaw Ulam realized that ENIAC could do such calculations much faster **using MC simulations**

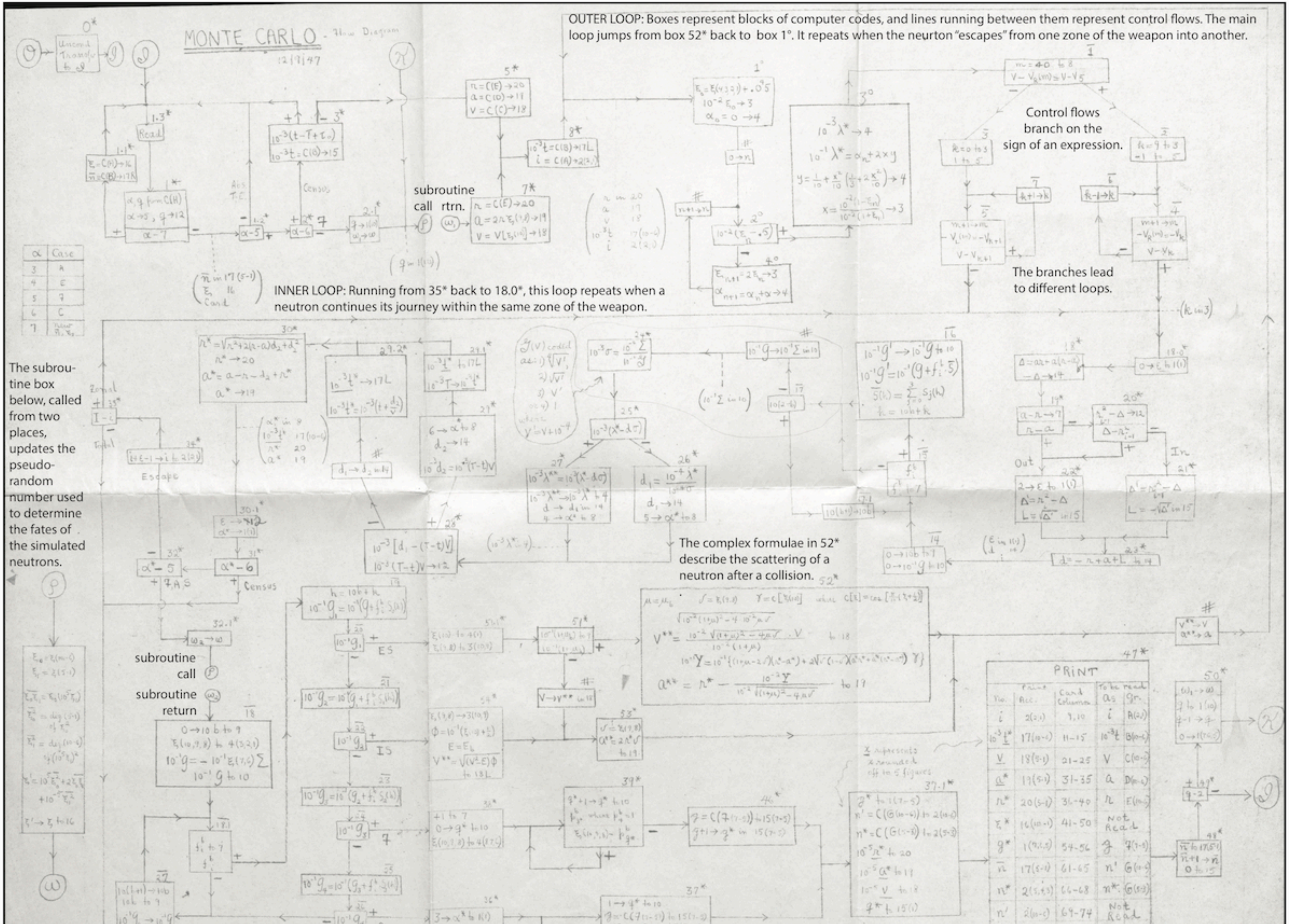
Stanislaw Ulam's brilliant idea

The first thoughts and attempts I made to practice [the Monte Carlo Method] were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations. Later [in 1946], I described the idea to John von Neumann, and we began to plan actual calculations.[13]



- Being secret, the work of them needed a code name
  - Nicholas Metropolis suggested using the name **Monte Carlo** which refers to the Monte Carlo Casino in Monaco, where Ulam's uncle would borrow money from relatives to go gamble
  - **Monte Carlo methods** became central for the simulations required for the Manhattan project and they became popular in other fields
    - Key ingredient: sequences of (pseudo)-random numbers

http://lib-www.lanl.gov/la-pubs/00326866.pdf



The program reads the details of a single simulated neutron from a punch card and traces its progress through the weapon. The "PRINT" block defines the data format of the output card holding updated information on that neutron.

The Monte Carlo method

A numerical technique for calculating probabilities and related quantities by using sequences of random numbers

Procedure (for the case of a single RV):

1. A series of random values $r_1, r_2, \ldots$ is generated according to a **uniform distribution** in the interval $0 < r < 1$

$$g(r) = \begin{cases} 1 & 0 < r < 1 \\ 0 & \text{otherwise} \end{cases}$$

Start here

2. The sequence $r_1, r_2 \ldots$ is used to determine another sequence $x_1, x_2 \ldots$ s.t. the x values are distributed according to a PDF $f(x)$ of interest

The values of x can then be treated as **simulated measurements**, and from them the *probabilities for x to take on values in a certain region can be estimated*

Uniformly distributed random numbers

Accomplished with algorithms called **random number generators**

One could in principle us a random physical process (e.g., tossing coins), but this is clearly not very practical

Simple example: Linear congruential algorithms (**LCG**)

- Starting from an an initial integer value n_0 (aka the *seed*)
  - generate a sequence of integers $n_1, n_2, \ldots$
  - according to $n_{i+1} = (a n_i) \mod m$

modulo operator: remainder of $\boldsymbol{a n_i}$ divided by $\boldsymbol{m}$

multiplier $\boldsymbol{a}$ and modulus $\boldsymbol{m}$ are integer constants

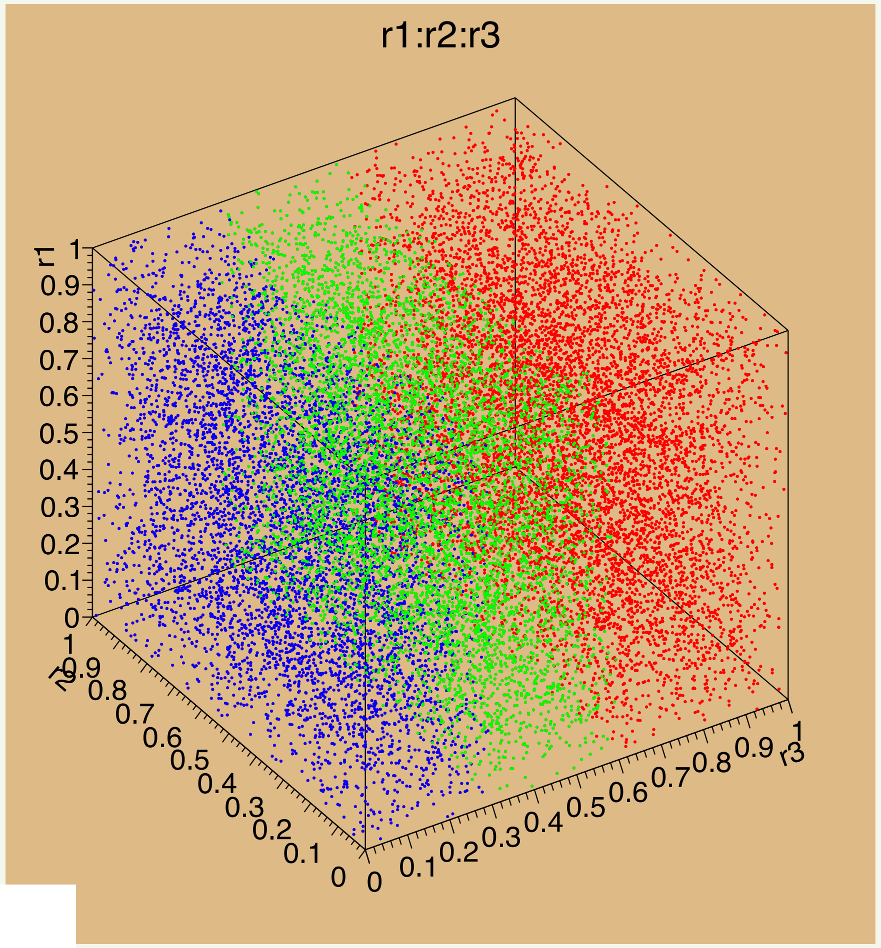
- n_i follow a periodic sequence in the range $[1,\ m-1]$
- $r_i = n_i / m$ are uniformly distributed in $[0,\ 1]$

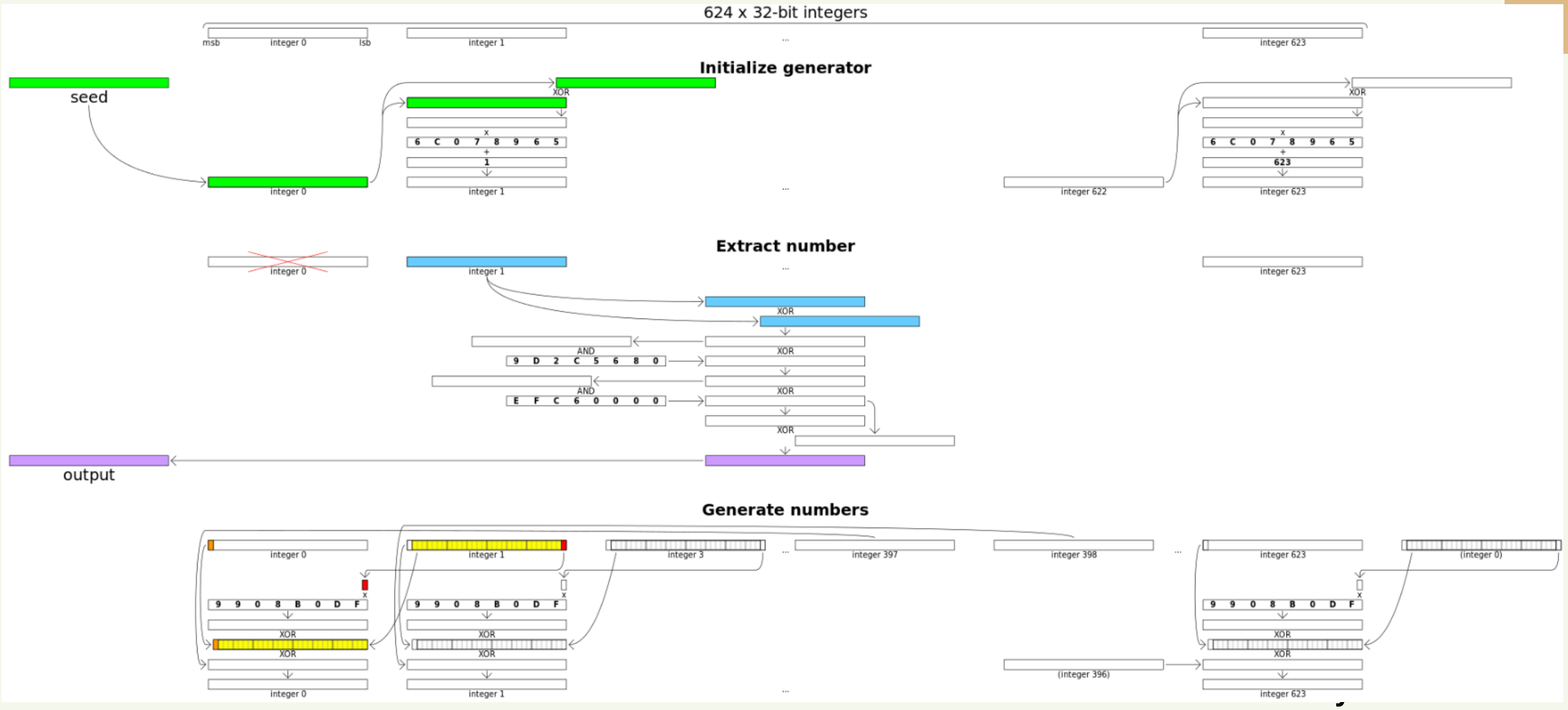
Sequence defined by $a,\ m,\ n_0$ so ***not truly random*** → *The resulting values are therefore called **pseudorandom***

Periods of 10^9 possible with well chosen values

Mersenne-Twister

- Based on Mersenne prime numbers
  - $(2^n - 1)$
  - Fully described by **624** integer numbers, which are used as **starting values**
    - With these calculate new values following an elaborated algorithm
- Extremely **long period** of about (2^{19937}-1 ≈ **10^{6000}**)
- Good distribution up to **623 dimensions** and performant





Visualization of generation of pseudo-random 32 bit integers using a Mersenne Twister Wikipedia

How random are random numbers?

- Several stringent tests exist
  - G. Marsaglia: **“Die-hard battery of tests of randomness”** (1995)
  - P.L. Ecuyer and R. Simard: **TestU01** (2007)
    - Small crush (10 Tests), Crush (96 Tests), Big Crush (106 Tests)

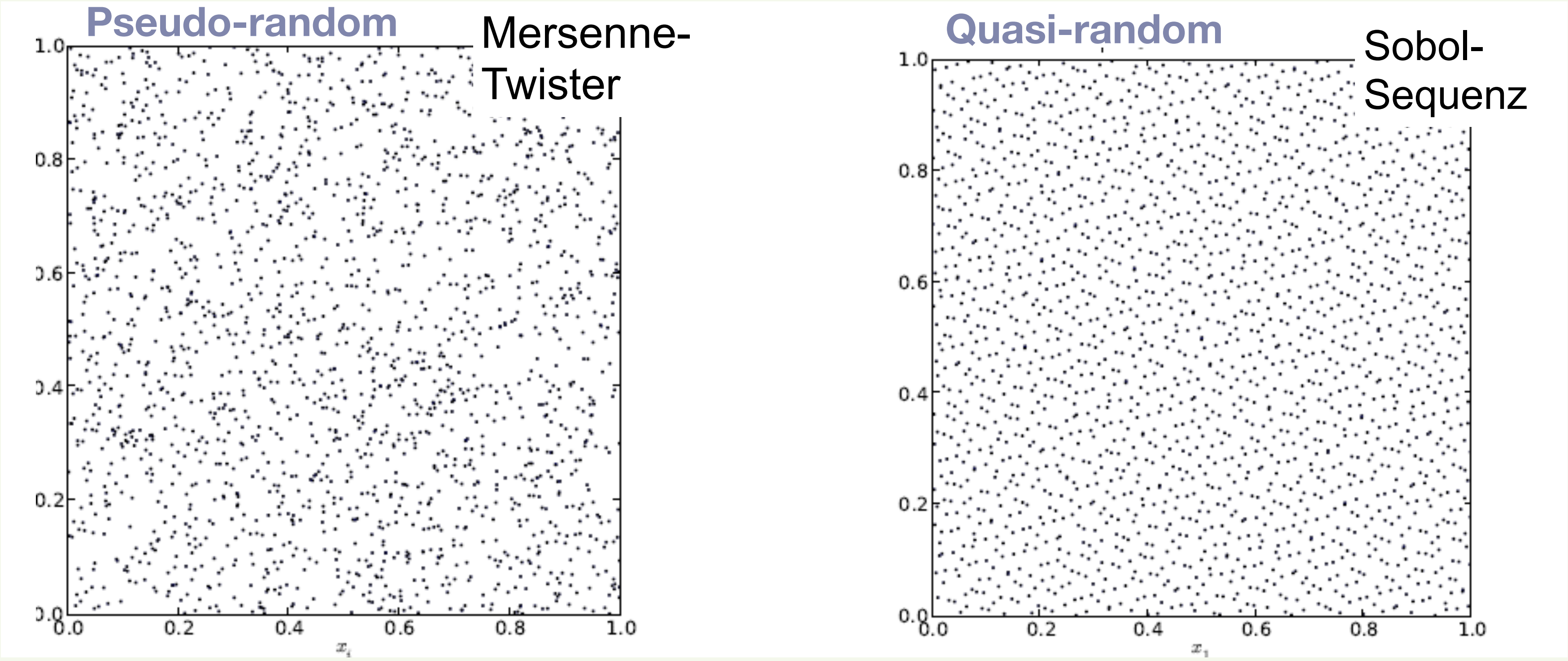
      http://dl.acm.org/citation.cfm?doid=1268776.1268777
- There exist more complicated generators that can pass the TestU01
- **Mersenne-Twister:**
  - does **not pass** all tests of ‘**Big Crush**’
  - **Disqualifies** its use in some applications, **e.g. cryptography**
  - But: For most applications you will encounter, Mersenne-Twister is fine

Try them (ROOT)

- `TRandom`: LCG
  - schnell, kurze Periode: 10^9
  - niedrigste Bits nicht unkorreliert, nicht verwenden!
- `TRandom1`: RANLUX (Lüscher, James ’94)
  - langsam, lange Periode: 10^{171}
  - übersteht TestU01 Suite (auf höchstem Level)
  - http://arxiv.org/abs/hep-lat/9309020
- `TRandom2`: Tausworthe (P.L’Ecuyer ’96)
  - schnell, Periodenlänge ok 10^{26}
- `TRandom3`: Mersenne-Twister (’98)
  - hinreichend schnell, Periodenlänge 10^{6000}
  - Default: `gRandom` points to `TRandom3`
- Methoden: Exp(tau), Integer(imax), Gaus(mean,sigma), Rndm(), Uniform(x1), Landau(mpv, sigma), Poisson(mean), Binomial(ntot, prob) u.v.m.

Alternative approach: Quasi-random #'s

- Pseudo-random numbers can get '**clots**':



http://web.maths.unsw.edu.au/~fkuo/sobol/joe-kuo-notes.pdf

https://en.wikipedia.org/wiki/Sobol_sequence

- Instead of aiming for a (pseudo)-randomness, try to get an equiprobable coverage of an n-dimensional space
  - This is possible with **Quasi-random numbers**
  - But: **Quasi-random numbers are correlated**, only use for integration applications (see later)

The Monte Carlo method (step 2)

A numerical technique for calculating probabilities and related quantities by using sequences of random numbers

Procedure (for the case of a single RV):

1. A series of random values $r_1, r_2 \ldots$ is generated according to a uniform distribution in the interval $0 < r < 1$

$$g(r) = \begin{cases} 1 & 0 < r < 1 \\ 0 & \text{otherwise} \end{cases}$$

2. The sequence $r_1, r_2 \ldots$ is used to determine another sequence $x_1, x_2 \ldots$ s.t. the x values are distributed according to a PDF $f(x)$ of interest

The values of x can then be treated as **simulated measurements**, and from them the *probabilities for x to take on values in a certain region can be estimated*

The transformation method

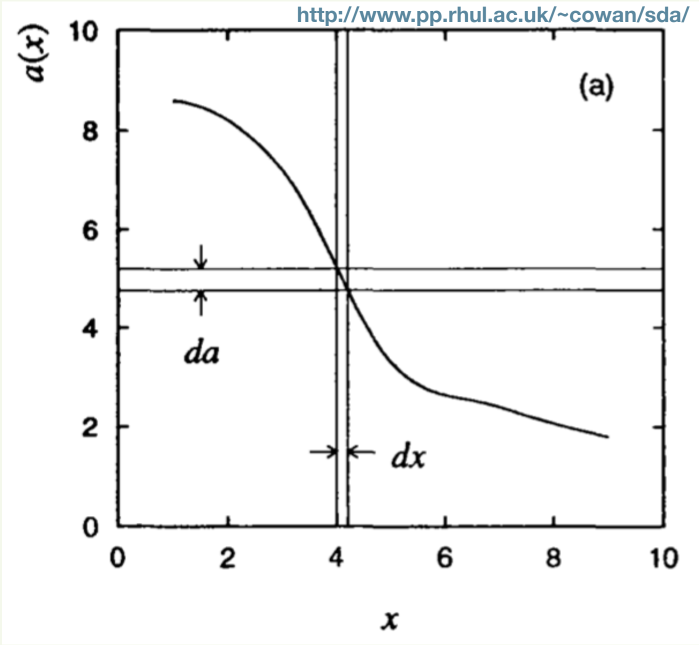
Given a sequence $r_1,\ r_2,\ \ldots$ uniformly distributed in $[0,1]$ step 1 ✓

Now → *Determine a sequence* $x_1,\ x_2,\ \ldots$ *distributed as PDF* $f(x)$

In the **transformation method** this is accomplished by finding a **suitable function** $\boldsymbol{x(r)}$ which directly yields the desired sequence *when evaluated with the uniformly generated* r *values*

Related to the transformation of variables described in Cowan sec. 1.4

$$g(a')da' = \int_{dS} f(x)dx$$



Here our task is to find a function $x(r)$ that is distributed according to a specified $f(x)$, given that r follows a uniform distribution between 0 and 1

In other words

Probability to obtain a value r in $[r, r+dr]$

$$g(r)dr = \int_{dS} f(x)dx$$

The probability to obtain a value of x in the corresponding interval $[x(r), x(r) + dx(r)]$

In order to obtain $x(r)$ s.t. this is true, one can require that:

Probability that $r < r'$ = Probability that $x < x(r')$

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

I.e., need to find a function $x(r)$ such that $G(r) = F(x(r))$

G and F are the CDFs corresponding to the PDFs g and f

Since $G(r) = r$ with $0 \leq r \leq 1$:

Uniform PDF [recall lecture 2, slides 61-62]

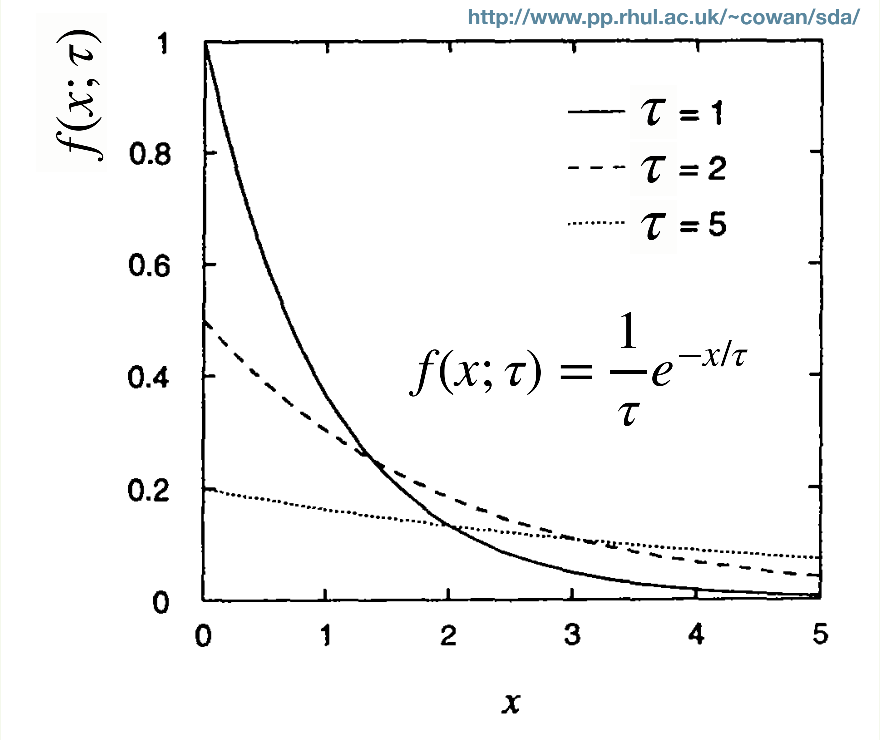
Depending on the $f(x)$ in question, it may or may not be possible to solve for $x(r)$ with

But first let's look at an example which we can solve analytically

Exponential distribution ...*as an example*

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

$$\int_0^{x(r)} \frac{1}{\tau} e^{-x'/\tau} dx' = r$$



Integrate and solve for x

$$x(r) = -\tau \log(1 - r)$$

Interpret: If r follows a uniform distribution between 0 and 1, then $x(r) = -\tau \log(1-r)$ will follow an exponential distribution

More examples of transformation functions

Dreieck-Verteilung

$$f(x) = 2x \quad 0 \leq x \leq 1$$
$$x(r) = \sqrt{r}$$

$$f(x) = (n+1)x^n \quad 0 \leq x \leq 1, n > -1$$
$$x(r) = r^{1/(n+1)}$$

Exponentialverteilung

$$f(x) = \gamma e^{-\gamma x}$$
$$x(r) = -\frac{1}{\gamma}\ln(1-r)$$

Breit-Wigner-Verteilung

$$f(x) = \frac{1}{\pi\Gamma/2}\frac{(\Gamma/2)^2}{x^2+(\Gamma/2)^2}$$
$$x(r) = -\frac{\Gamma}{2}\tan\left[\pi(r-\frac{1}{2})\right]$$

Log-Weibull-Verteilung

$$f(x) = e^{-x-e^{-x}}$$
$$x(r) = -\ln(-\ln r)$$

Paar von Gauß-Zahlen

$$f(x,y) = \frac{1}{2\pi}\exp\left[-\frac{x^2+y^2}{2}\right]$$
$$x(r_1,r_2) = \sqrt{2\ln(r_1-1)}\cos(2\pi r_2)$$
$$y(r_1,r_2) = \sqrt{2\ln(r_1-1)}\sin(2\pi r_2)$$

Anmerkung:
r kann auch durch 1-r ersetzt werden

z.B. Bohm/Zech Abschnitt 4.2

Acceptance-rejection method

The distribution of the scattering angle θ in the reaction $e^+e^- \to \mu^+\mu^-$ with $x = \cos\theta$

$$f(x) = \frac{3}{8}(1 + x^2) \qquad -1 \leq x \leq 1$$

However, as we just said, it is often too difficult to solve for $x(r)$ analytically, ***so…***

1) Generate a random number x, uniformly distributed between x_{min} and x_{max}

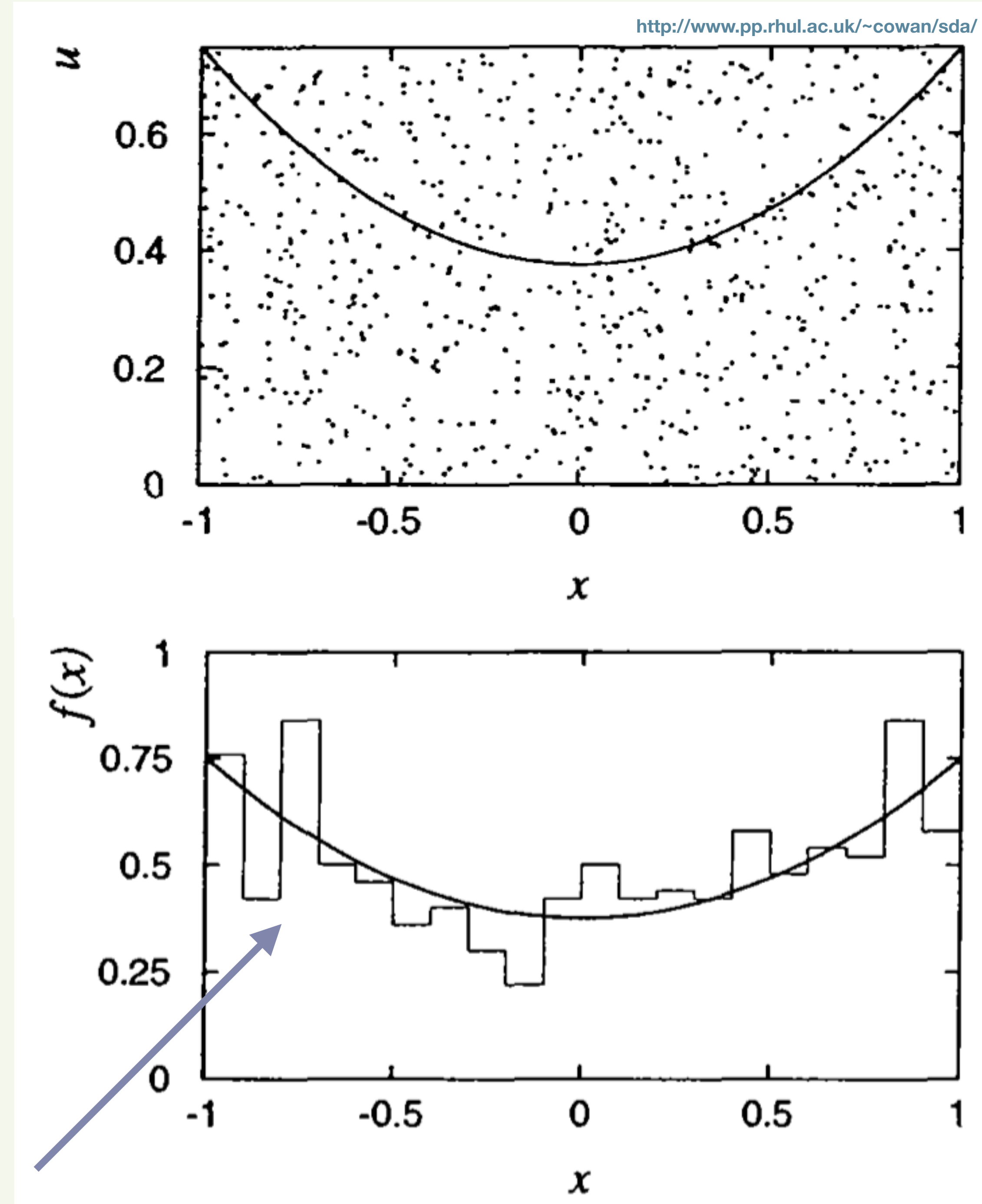
$$x = x_{min} + r_1(x_{max} - x_{min})$$

(r_1 is uniformly distributed between 0 and 1)

2) Generate a second independent random number u uniformly distributed between 0 and f_{max}.

$$u = r_2 f_{max}$$

3) If $u < f(x)$, then accept x. If not, reject x and repeat

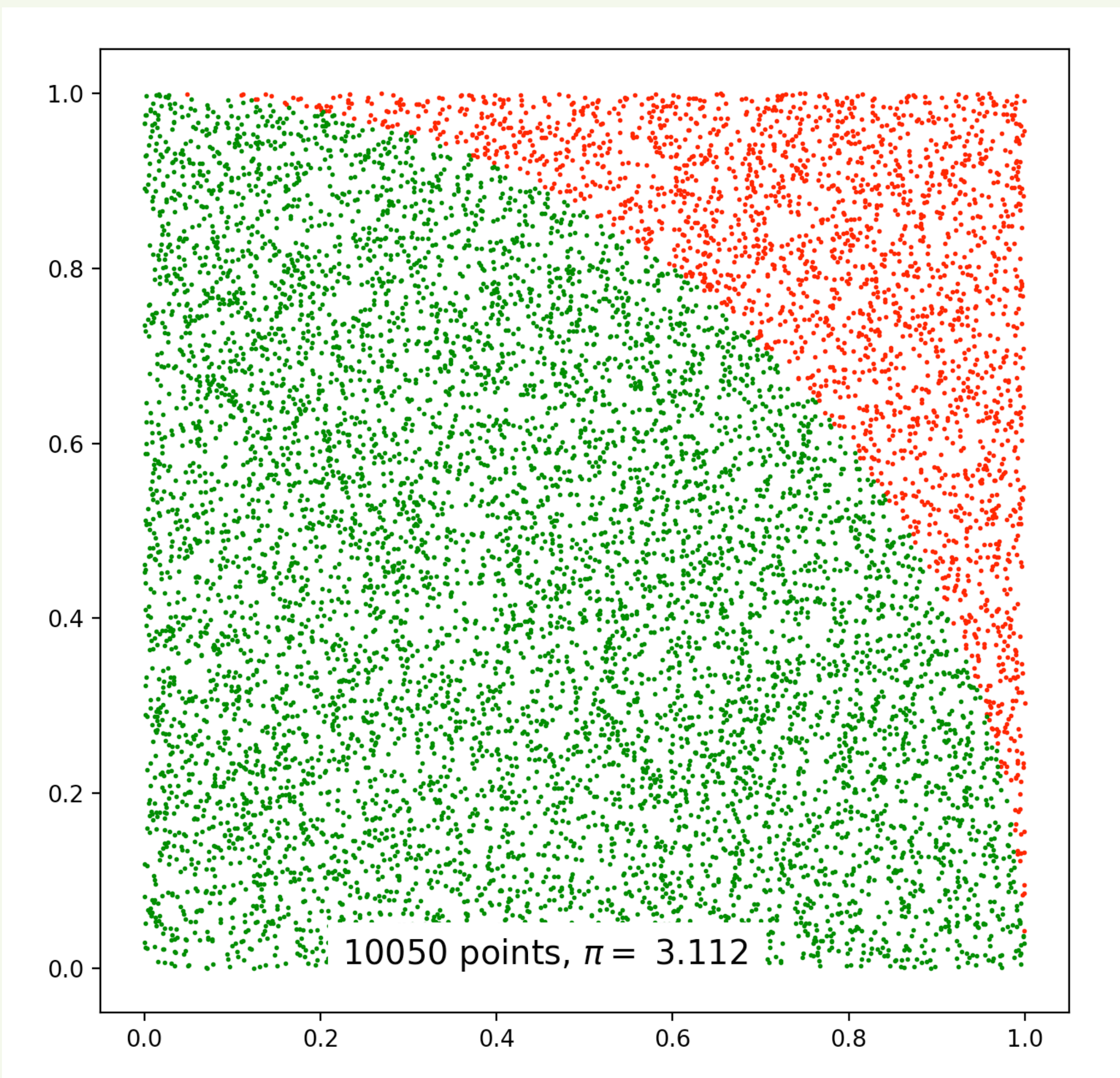


A normalized histogram constructed from the accepted points

How accurate is MC integration?

Ratio of accepted and rejected+accepted values proportional to $\int_S f(x)dx$

$\Rightarrow$ ***MC Integration***

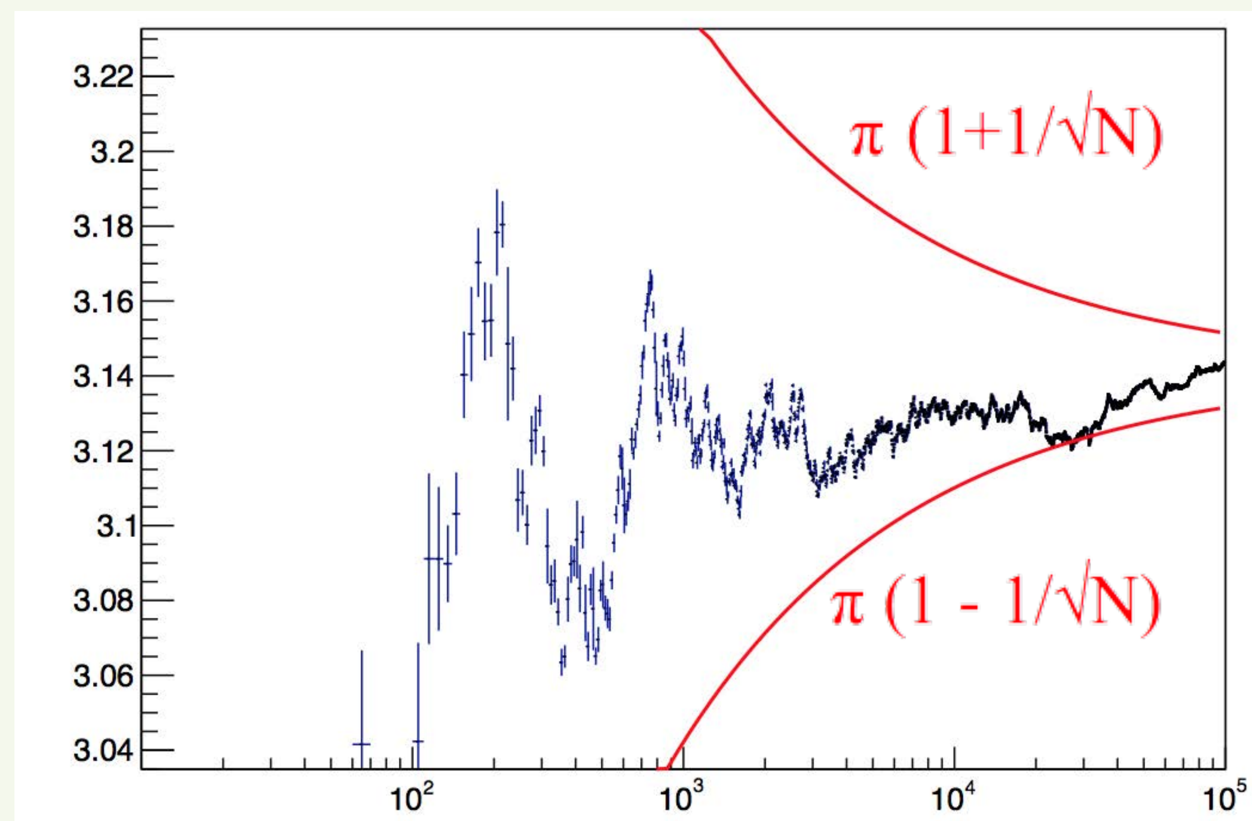


$$\frac{N_a}{N_a + N_r} \qquad N = N_a + N_r$$

naive error propagation assuming Gaussian ($\sqrt{N}$) errors

$$\sqrt{\frac{N_a^2 N_r}{(N_a + N_r)^4} + N_a\left(-\frac{N_a}{(N_a + N_r)^2} + \frac{1}{N_a + N_r}\right)^2} \approx \sqrt{\frac{1}{N_a + N_r}} = \sqrt{\frac{1}{N}}$$

dropping all higher order terms and use $N_a + N_r \approx N_a$ for large N



Converges as $\sim 1/\sqrt{N}$ for large N

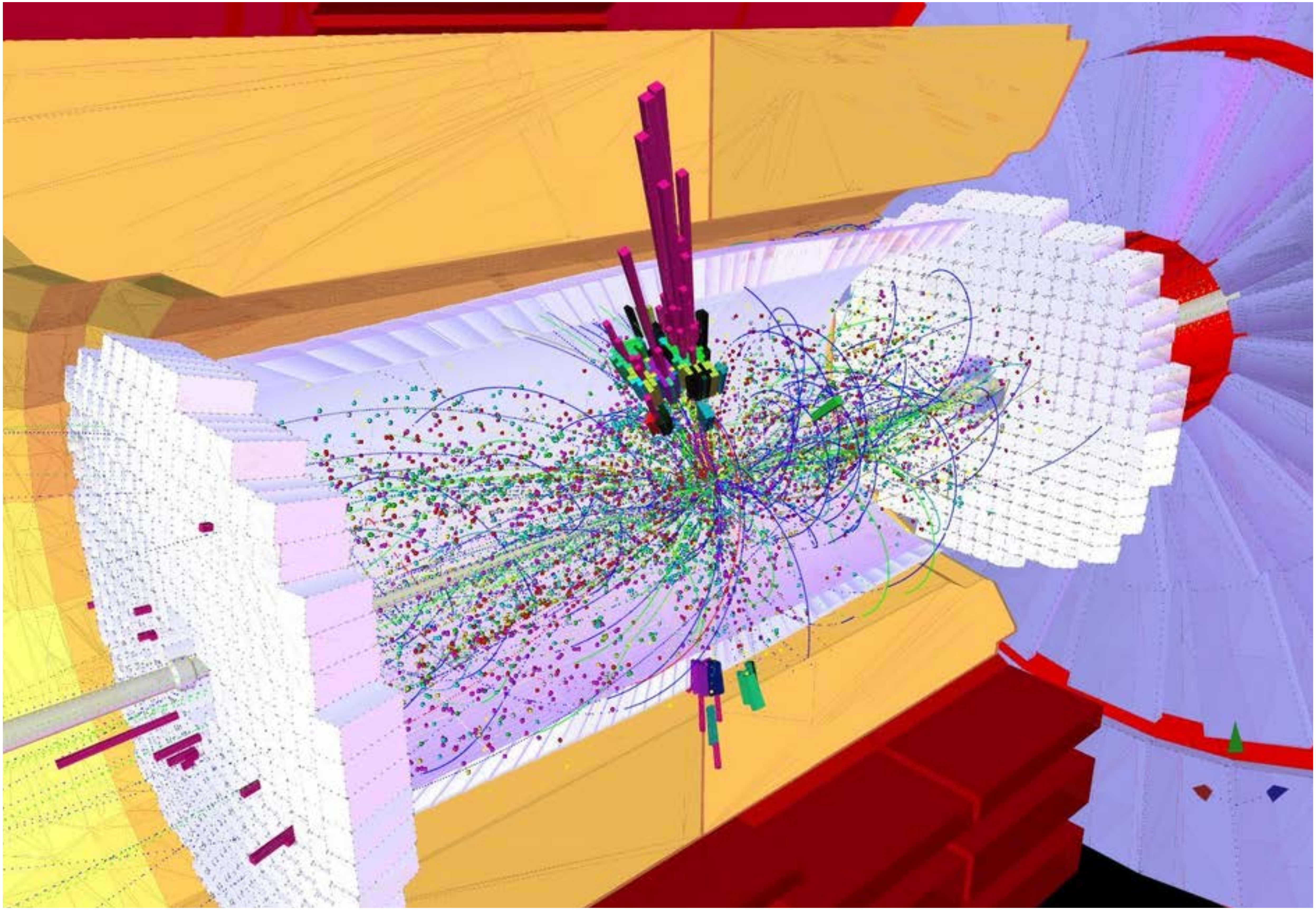
Other methods to increase performance

- If integration is the goal:
  - Subtract known analytical function $f(x)$ that approximates $g(x)$
    - Can use analytical expression for part of the integral

$$\int_a^b g(x)dx = \int_a^b f(x)dx + \int_a^b |g(x) - f(x)|\,dx$$

  - Stratified sampling
    - Increase sampling rate in regions of interest to increase precision there (e.g. in tails of distributions if such are relevant for what you are doing)

Examples of MC in HEP



Application examples:

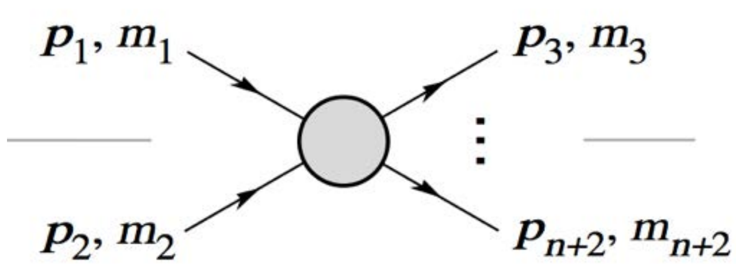
- Calculation of a collision cross section
  - Need to calculate matrix elements, flow factors and phase space

Matrixelement

$$d\hat{\sigma} = \frac{(2\pi)^4|\mathcal{M}|^2}{4\sqrt{(p_1p_2)^2 - m_1^2m_2^2}} \times d\Phi_n(p_1+p_2;p_3...p_{n+2})$$

Flussfaktor

Phasenraum



  - Multidimensional integral, often no analytical calculation possible
  - In case of hadron collisions (e.g. LHC): need to integrate over parton density functions that describe the momentum distribution of quarks inside a proton (or anti-proton)

$$\sum_{ij}\int dx_1dx_2f_i(x_1,Q^2)f_j(x_2,Q^2)\hat{\sigma}(Q^2)$$

Application examples

- Finite resolution of a detector can be described via a folding integral:
  - Instead of true value x, one measures a smeared out value x'
    - The resolution can be described via a function $t(x, x')$ and a given distribution $f(x)$ is measured as

$$f'(x') = \int_{-\infty}^{\infty} t(x, x')f(x)dx$$

  - In practice measured distributions are smeared out by a large number of individual processes
    - Intrinsic resolution of the detector, electronic noise, digitization fragments, and other systematic shifts. The integral is not one-dimensional but a multi-dimensional entity
    - Can simulate such effects one-by-one in MC simulations and the calculation of this integral becomes to the task of adding up random variables.

For next time

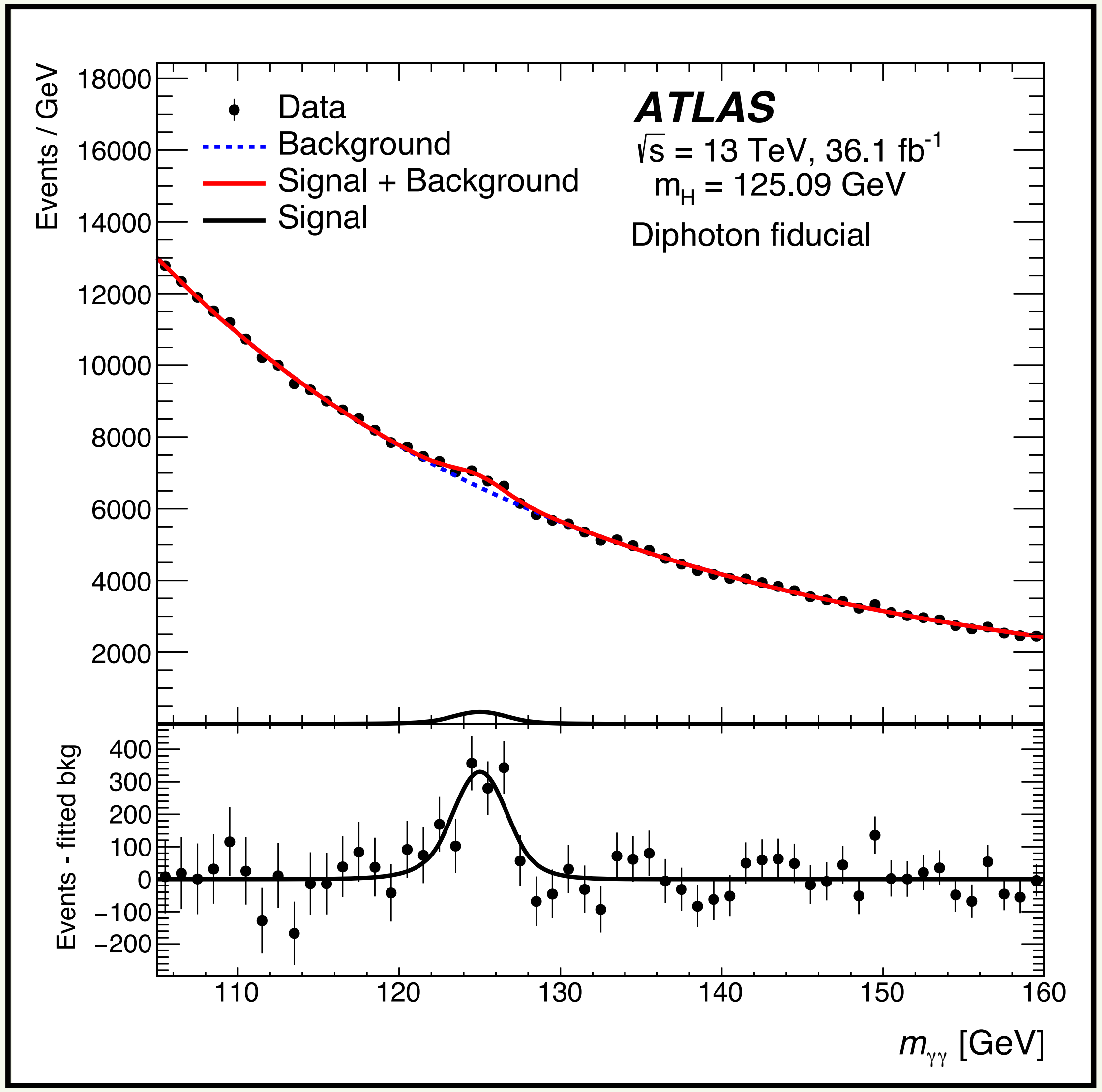
- Required reading
  - Cowan textbook: chapters 2 & 3
  - Lista textbook: chapter 4
  - Reading material / L03 / WhyTheNormalDistribution?

- Extra reading for fun: /Reading material / L03 /
  - InfoTheoryAndMaxEntropy
  - IntroQuasiRandomNumbers
  - TheBeginningOfTheMCMethod
  - TheEvolutionOfTheNormalDistribution

} *For fans of history*

Next time

- Introduction to parameter estimation
  - General concept of parameter estimation
  - Introduction into the method of Maximum Likelihood



Quiz Time: 3rd Round

MC method

- How would you write an algorithm to integrate a multi-dimensional function (of dimensionality N) using the 'acceptance-rejection' method?
- Address in particular: what ingredients you need and sketch out the algorithm explicitly.

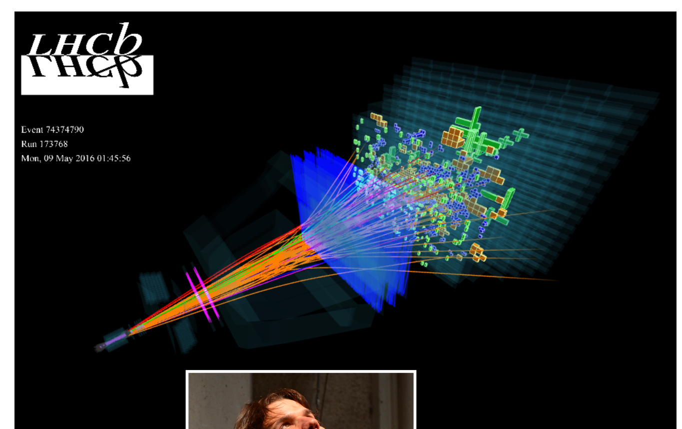
KCETA Colloquium

Exploring new horizons with flavour at the LHCb experiment

Thursday, May 11, 2023
Kleiner Hörsaal A (CS) 15:45 - 17:00

Vitalii Lisovskyi
(École Polytechnique Fédérale de Lausanne (EPFL))

The recent decade has seen rapid developments in heavy-flavour physics. Among others, the LHCb experiment has delivered a number of important results. Precision studies of beauty and charm hadrons, their properties and decays, have not only improved our understanding of the flavour structure of the Standard Model, but also revealed a number of intriguing anomalies. This talk will present highlights from the LHCb experiment, highlighting the status of the recent anomalies in heavy-flavour decays.



Please note:
The colloquium will also be live-streamed to B401 SR 410 (CN).

KIT Center Elementary Particle and Astroparticle Physics (KCETA)
www.kceta.kit.edu

Karlsruher Institut für Technologie

Bibliography

- Part of the material presented in this lecture is adapted from the following sources. See the active links (when available) for a complete reference
  - **Probability for CS** (Stanford): slides 4-6, 9-26,
  - **Statistical Data Analysis** textbook by G. Cowan (U. London): images on slides 8, 33-37, 53, 55, 57, 59