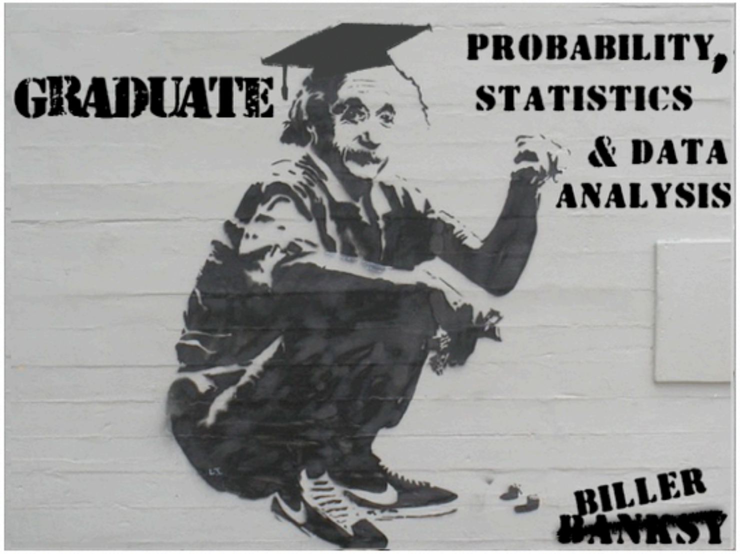
Home Research Teaching Outreach Physics Poems Contact

http://www-pnp.physics.ox.ac.uk/~biller/grad-stats-course.html



Problem Set

Likelihood Exercise

Lecture 1
Lecture 2
Lecture 3
Lecture 4
Lecture 5
Lecture 6
Lecture 7
Lecture 8
Lecture 9
Lecture 10
Lecture 11
Lecture 12

SUGGESTED BOOKS

"Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods", Behnke *et al.* (2013)"

Statistical Data Analysis", G. Cowan (1998)

"Numerical Recipes", W. Press, S. Teukolsky, W. Vettering & B. Flannery (2007)

"A Guide to the Use of Statistical Methods in the Physical Sciences", R. J. Barlow (2008)

"Data Analysis: A Bayesian Tutorial" by D. Sivia (1996)

"Measurements and their Uncertainties: A practical guide to modern error analysis" by I. Hughes and T. Hase (2010)

"A Student's Guide to Bayesian Statistics" by Ben Lambert (2018)

"Probability Theory: An Introductory Course" by Ya. G. Sinai (1992)

OVERVIEW

Lecture 1: Probability traps; Binomial & Poisson Distributons; Expectation and Variance; Estimators; Gaussian Distribution

Lecture 2: Central Limit Theorem; Properties of Normal Distributions; Look elsewhere effect; Regression to the Mean; Correlations

Lecture 3: Uncertainties & Error Propagation; Testing Models: chi-squared; "Scientific Method"

Lecture 4: Student's t; Correlation test; Non-parametric tests: rank and K-S; What is 'Normal?'; Robust parameter estimation

Lecture 5: p-values; Combined p-values as a Statistic; Maximum Likelihood; Neyman-Pearson Lemma; Wilks' Theorem; Joint analysis of multiple data sets; Extended Likelihood; Nuisance parameters and profile likelihood

Lecture 6: Asimov data sets; Propagating PDF uncertainties; Bayes' Theorem

Lecture 7: *Confidence and Credibility intervals*: Asking the right question; Wilks (again); Neyman construction; "Standard" and "Feldman-Cousins," Interpretation issues; Bayesian intervals and Priors

Lecture 8: *Loose Ends:* Real ensembles; Constant priors; Robust data comparisons; *Data Presentation:* 'Binsmanship' and Dodgy Error Bars; More Things to Avoid; Ways to Display Uncertainties; Visualising Multi-Dimensional Data; Boxes, Whiskers and Violins

Lecture 9: Blind Analysis; Bifurcated side-band analysis, Data "Correction"; Statistical Optimisation; Redundancy

Lecture 10: *Monte Carlo methods:* Distribution sampling; Markov chains; MC Integration; Smart sampling, Weighted sampling

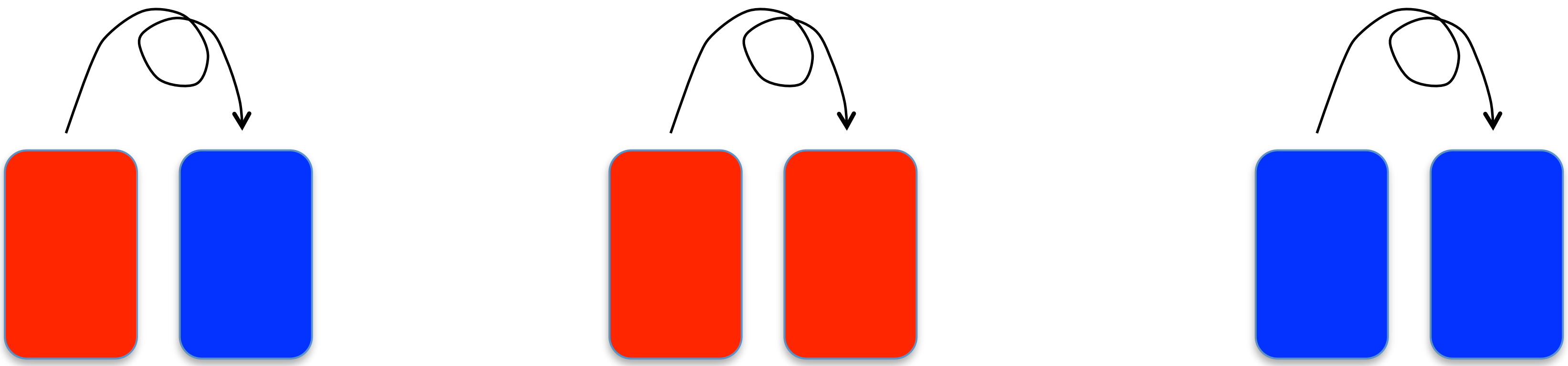
Lecture 11: *Optimisaton Methods:* Grid Search; Golden Ratio; Powell's Method, Gradient Descent; Determination of posterior densities via Markov Chain Monte Carlos

Lecture 12: Fisher Linear Discriminant, Decision trees & boosted decision trees

Lecture 1:

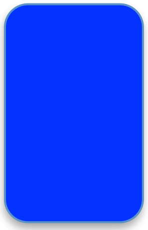
- Context
- Probability traps
- Binomial & Poisson Distributions
- Expectation and Variance
- Estimators
- Gaussian Distribution

“The Card Game”



“The Card Game”

I’ll bet you £10 that the other side is blue.



side shown	other side
1 (R,R,B)	B
2 (B,R,B)	R,B

Chance for the other side to be blue is 2/3 !

“Prisoner’s Paradox”

One of you lucky boys
will **<u>only</u>** get life in prison. But
I have instructed the guard not to
inform you whether or not you will
hang until I announce to the press
tomorrow morning as a
last minute surprise!

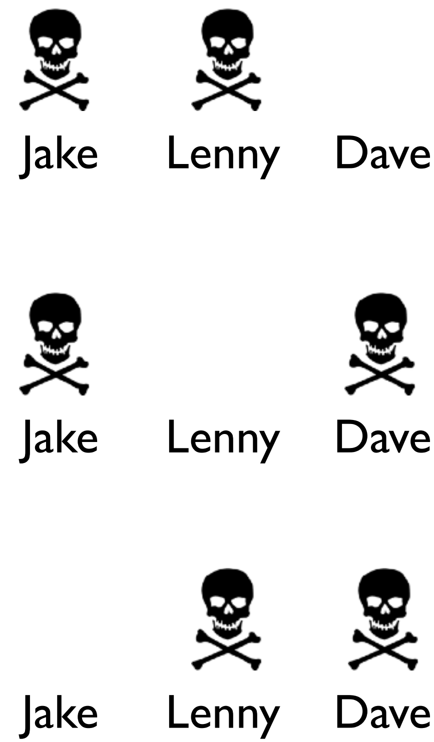
Not good!

Survival probability = 1/3 !!

Jake Lenny Dave

“Prisoner’s Paradox”

“Prisoner’s Paradox”



S S F S F S

“Bernoulli Trials”

Loaded Dice?

If not loaded (test to reject H0),

$$1/6 \times 1/6 \times 1/6 \times 1/6 \times 1/6 \times 1/6 = \frac{1}{46656}$$

What’s the chance probability of getting **exactly** this sequence?

$$1/6 \times 1/6 \times 5/6 \times 1/6 \times 5/6 \times 1/6 = \frac{25}{46656}$$

or more generally: $p^k (1-p)^{n-k}$

where p is the probability of successes (getting a 3), k is the number of successes and n is the total number in the sequence.

What’s the chance probability of getting four 3’s in this order?

S S F S F S

$(1/6 \times 1/6 \times 5/6 \times 1/6 \times 5/6 \times 1/6)$(any 4 of 6)

distinct ways to order 6 things

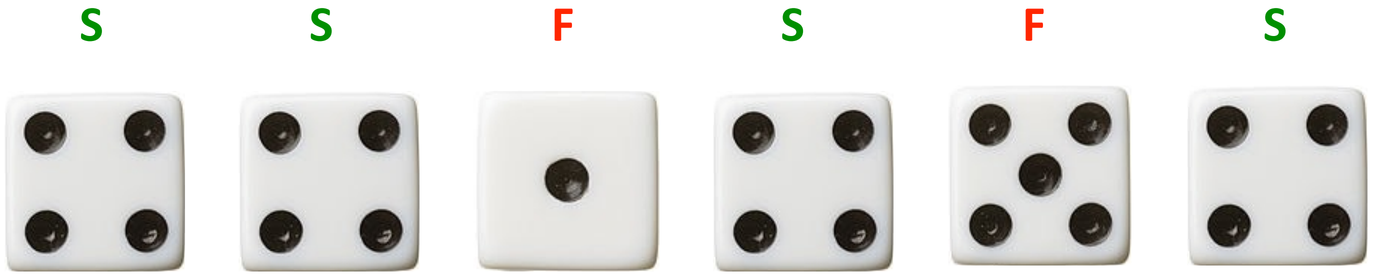
$$\text{(any 4 of 6)} = \binom{6}{4} = \frac{6!}{4!2!} = 15 \rightarrow \frac{375}{46656}$$

re-ordered of successes is not distinct

re-ordering of failures is not distinct

or more generally: $\boxed{\binom{n}{k} p^k (1-p)^{n-k}}$ Binomial Distribution ("two terms")

What's the chance probability of getting four 3's in any order?

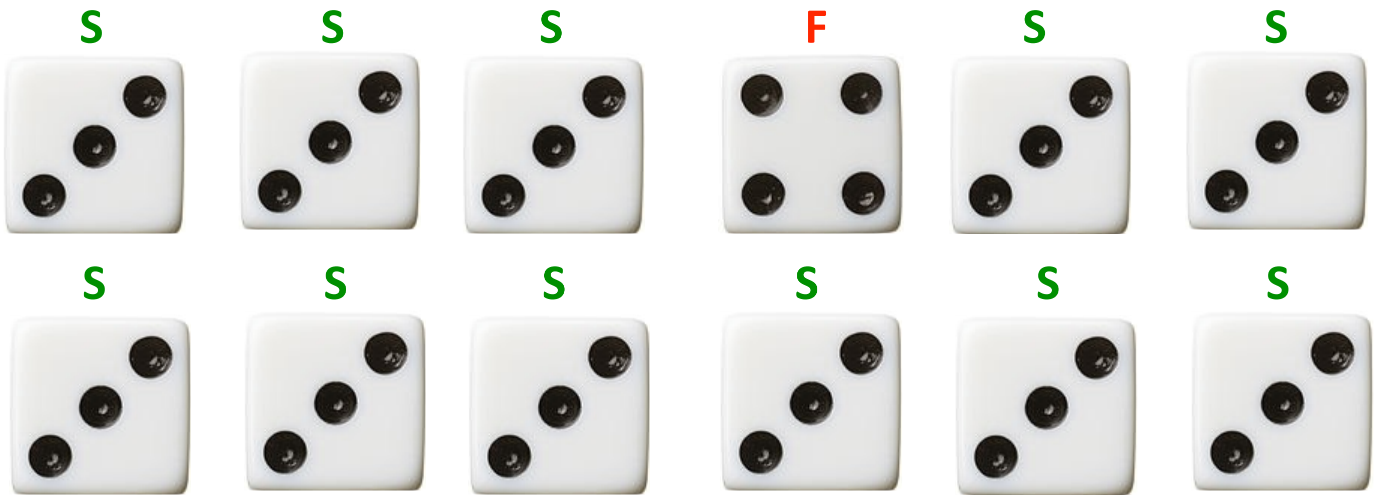


what about this?

so we really want $6 \text{ x } \binom{6}{4} (1/6)^4 (1-1/6)^2$

$$= \frac{2250}{46656}$$

What's the chance probability of getting four of anything in any order?



But these would draw at least as much suspicion!

so, in this case we want

$$\sum_{j=k}^{n} 6x\binom{n}{j}\, p^{j}\,(1-p)^{n-j}$$

k=4, n=6

$$\frac{2250}{46656} + \frac{180}{46656} + \frac{6}{46656} = \frac{2436}{46656} = 5.2\%$$

What's the chance probability of getting four or more of any number in any order?

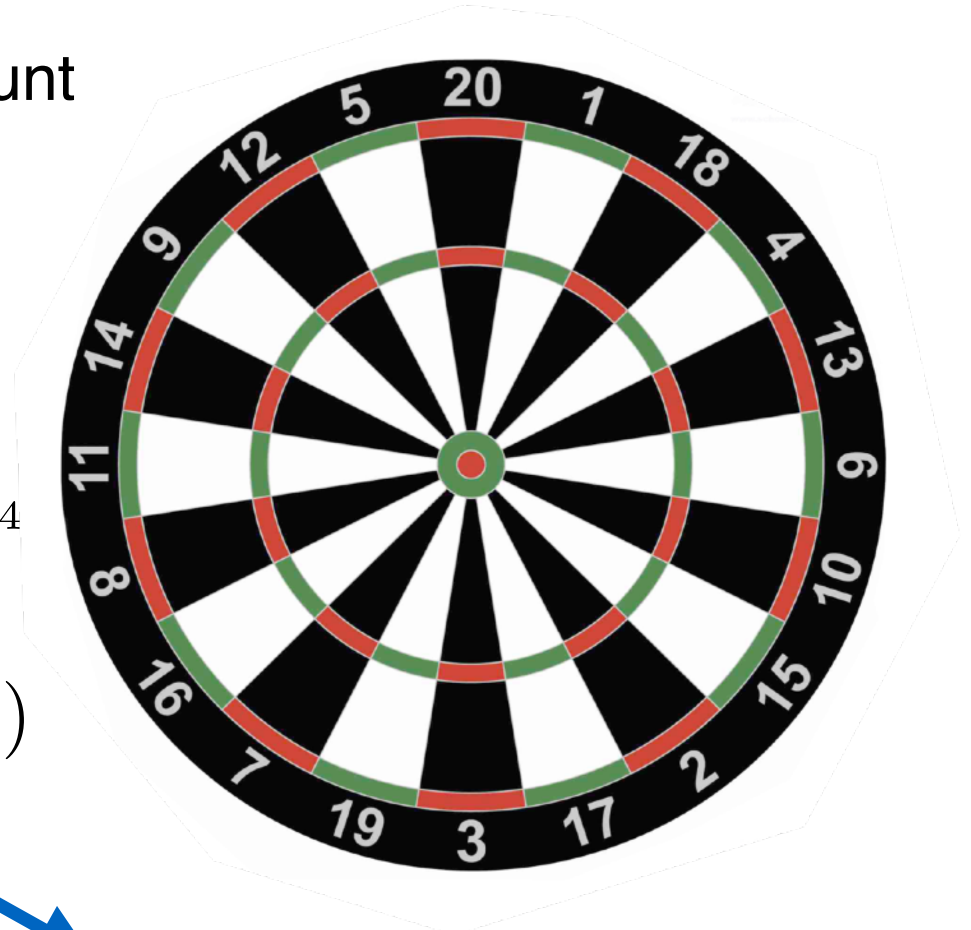
Statistical probability is basically the frequency with which a given “equivalent” outcome occurs if we were to repeat the same experiment over and over again.

What is the source of this statistical behaviour??

1) Hidden variations in initial conditions

2) Fundamental uncertainty (quantum mechanics)

Assume terrible aim, but only count throws that hit dart board. . .



What's the chance of hitting the bullseye given 100 throws?

$$p_s = (0.5in/17.75in)^2 = 7.93 \times 10^{-4}$$

$$P_{tot} = \sum_{k=1}^{100} P_{bin}(\text{k successes})$$

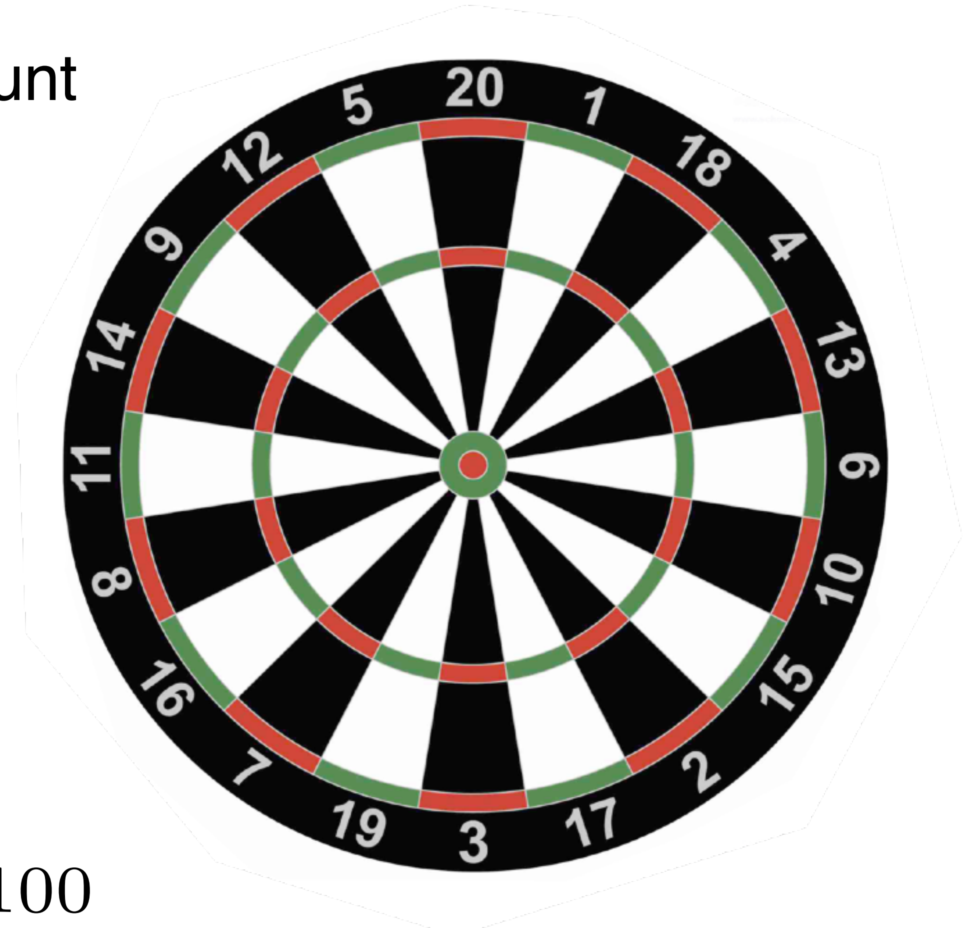
$$\left(\frac{100!}{k!(100-k)!}\right) p_s^k (1-p_s)^{100-k}$$

$$= 1 - P_{bin}(0 \text{ successes})$$

$$= 1 - (1-p_s)^{100}$$

$$= 7.63\% \qquad \sim 100 \times p_s$$

Assume terrible aim, but only count throws that hit dart board. . .



What's the chance of hitting the **20** given 100 throws?

$$p_s \sim 1/20 = 0.05$$

$$P_{tot} = 1 - (1 - 0.05)^{100}$$

$$= 99.4\% \qquad \neq 100 \times 0.05 \text{ !!!}$$

Binomial Distribution:

$$P(\text{k successes in n attempts}) = \left(\frac{n!}{k!(n-k)!}\right) p_s^k (1 - p_s)^{n-k}$$

prob of each success

So, the expected (average) number of successes after summing over **n** identical Bernoulli trials is:

$$\mu = np$$

Now consider the case where the expected number of successes depends on the size of a **continuous variable** (*e.g.* length or time interval), which can be arbitrarily small.

The number of successes expected over a continuous interval of finite size can be viewed as resulting from the sum of an infinite number of Bernoulli trials carried out for arbitrarily small intervals such that:

$$\mu = \lim_{n \to \infty} np$$

So, set p= μ/n and evaluate

$$P(k) = \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k}$$

$$= \left(\frac{\mu^k}{k!}\right) \lim_{n\to\infty} \boxed{\frac{n!}{(n-k)!} \left(\frac{1}{n}\right)^k} \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-k}$$

$$\lim_{n\to\infty} \frac{n!}{(n-k)!} \left(\frac{1}{n}\right)^k = \lim_{n\to\infty} \frac{n(n-1)(n-2)...(n-k)(n-k-1)...(1)}{(n-k)(n-k-1)...(1)} \left(\frac{1}{n}\right)^k$$

$$= \lim_{n\to\infty} \frac{n(n-1)(n-2)...(n-k+1)}{n^k}$$

$$= \lim_{n\to\infty} \left(\frac{n}{n}\right) \left(\frac{n-1}{n}\right) \left(\frac{n-2}{n}\right) ... \left(\frac{n-k+1}{n}\right)$$

$$= 1$$

So, set p= μ/n and evaluate

$$P(k) = \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k}$$

$$= \left(\frac{\mu^k}{k!}\right) \lim_{n\to\infty} \underbrace{\boxed{\left(1 - \frac{\mu}{n}\right)^n}} \left(1 - \frac{\mu}{n}\right)^{-k}$$

$$\lim_{n\to\infty} \left(1 - \frac{\mu}{n}\right)^n = \lim_{n\to\infty} \exp\left[\log\left(1 - \frac{\mu}{n}\right)^n\right]$$

$$= \lim_{n\to\infty} \exp\left[n \log\left(1 - \frac{\mu}{n}\right)\right]$$

$$= \exp\left[n\left(-\frac{\mu}{n}\right)\right]$$

$$= e^{-\mu}$$

So, set p= μ/n and evaluate

$$P(k) = \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k}$$

$$= \left(\frac{\mu^k}{k!}\right) \lim_{n\to\infty} \qquad e^{-\mu} \qquad \boxed{\left(1 - \frac{\mu}{n}\right)^{-k}}$$

$$\lim_{n\to\infty} \left(1 - \frac{\mu}{n}\right)^{-k} = 1$$

So, set p= μ/n and evaluate

$$P(k) = \lim_{n\to\infty} \frac{n!}{k!(n-k)!}\left(\frac{\mu}{n}\right)^k \left(1-\frac{\mu}{n}\right)^{n-k}$$

$$= \left(\frac{\mu^k}{k!}\right) \lim_{n\to\infty} \qquad\qquad e^{-\mu}$$

$$= \boxed{\frac{\mu^k e^{-\mu}}{k!} \quad \text{Poisson Distribution}}$$

Counting statistics, decay processes… ➡ continuous variable is time

Interaction lengths ➡ continuous variable is distance

Radioactive Decay:

1 or more!

What's the probability of detecting a decay from a radioactive source after some time t ?

τ = average time for a decay to occur (mean lifetime)

μ = average # decays in time t, which must be t/τ

Probability for no decays (n=0) within time t

$$P_0 = \left(\frac{\mu^n e^{-\mu}}{n!}\right) \longrightarrow e^{-t/\tau}$$

$$\longrightarrow \quad P_{decay} = 1 - e^{-t/\tau}$$

(integrated over the time interval)

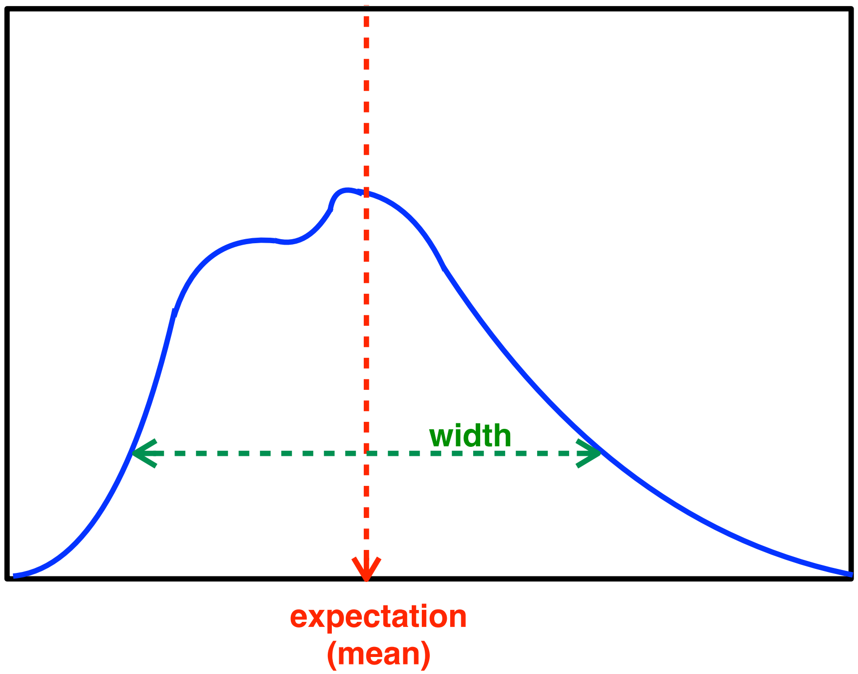
Differential Probability: $$P'(t) = \frac{1}{\tau} e^{-t/\tau}$$

Note that this is now a probability for a continuous quantity!

Poisson distribution: the probability of success depends on <u>continuous</u> variable (μ), but the observation is a <u>discreet</u> number of successes (n).

But observations are not always of a discreet variable. For continuous random variables (*i.e.* time, length, *etc.*), the probability of obtaining a particular **exact** value is generally vanishingly small (no phase space!). But the relative probability of getting a value in this vicinity versus that vicinity is meaningful. That's when you talk about "probability densities".

But the terms "probability distribution" and "probability density function" are sometimes informally used interchangeably.


width
expectation
(mean)

Variance: "Average Squared Deviation from Mean"

note: $$\boxed{\langle (x-\mu)^2 \rangle} = \langle x^2 \rangle + \mu^2 - 2\mu\langle x \rangle \boxed{= \langle x^2 \rangle - \mu^2}$$

for Poisson:

$$\langle n^2 \rangle = \sum_{n=0}^{\infty} n^2 \frac{\mu^n}{n!} e^{-\mu} = e^{-\mu} \sum_{n=1}^{\infty} n \frac{\mu^n}{(n-1)!}$$

$$= e^{-\mu} \sum_{n=1}^{\infty} \left[(n-1)\frac{\mu^n}{(n-1)!} + \frac{\mu^n}{(n-1)!} \right] = e^{-\mu} \left[\sum_{n=2}^{\infty} \frac{\mu^n}{(n-2)!} + \sum_{n=1}^{\infty} \frac{\mu^n}{(n-1)!} \right]$$

$$= e^{-\mu} \left[\mu^2 \sum_{n=2}^{\infty} \frac{\mu^{n-2}}{(n-2)!} + \mu \sum_{n=1}^{\infty} \frac{\mu^{n-1}}{(n-1)!} \right] = e^{-\mu} \left[\mu^2 (e^{\mu}) + \mu (e^{\mu}) \right]$$

$$= \mu^2 + \mu$$

$$\boxed{\sigma^2 =} \langle (n-\mu)^2 \rangle = \langle n^2 \rangle - \mu^2 = \boxed{\mu}$$

variance = $\sigma^2 = \langle x^2 \rangle - \mu^2$

Units of σ are same as units of x (or μ)

But, for Poisson, $\sigma^2 = \mu$ How do units work?

Here, μ refers to the expected number of successes, which is unit-less (special case)

$$\sigma = \sqrt{\langle (x-\mu)^2 \rangle} = \sqrt{\langle x^2 \rangle - \mu^2}$$

= “RMS (Root Mean Squared) deviation”

universal

“Standard deviation”

when interpreted in the context of a Normal (Gaussian) distribution

Some Useful Consequences:

- The RMS deviation on a measured number of counts due to statistical fluctuations is the square root of the expected mean number of counts (sqrt of the measured number is often not a bad approximation)
- For a large numbers of events, the expected sensitivity for detecting a signal in a counting experiment in terms of the number of standard deviations above background fluctuations is $\sim S/\sqrt{B}$
- In a counting experiment, the number of signal and background events detected are proportional to the counting time. Thus, the signal sensitivity goes like $\sqrt{T}$ in the large n limit

Variance in the Estimated Mean

Note that: $\mathrm{var}(\alpha x) = \left\langle (\alpha x)^2 \right\rangle - \langle \alpha x \rangle^2 \quad = \alpha^2 \left(\left\langle x^2 \right\rangle - \langle x \rangle^2 \right)$

$$= \alpha^2 \mathrm{var}(x)$$

So, consider: $\sigma_m^2 = \mathrm{var}\left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) = \frac{1}{n^2} \mathrm{var}\left(\sum_{i=1}^{n} x_i \right)$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{var}\left(x_i\right)$$

For independent variables (as will be shown in lecture 4)

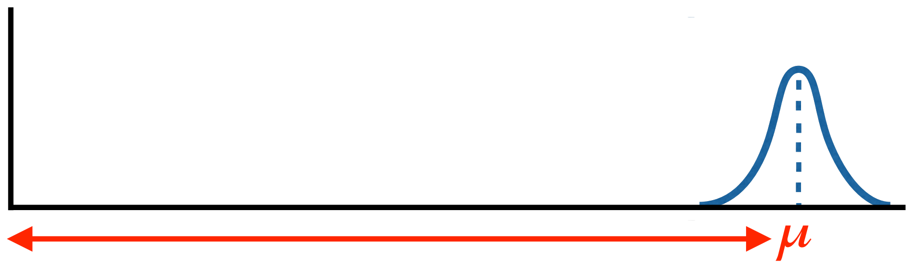
$$= \frac{1}{n^2}\left(n\sigma^2\right) \quad = \frac{\sigma^2}{n}$$

or

$$\sigma_m = \frac{\sigma}{\sqrt{n}}$$

Gaussian (Normal) Distribution as a Limiting Case of Poisson Statistics

Assume μ and n large, with n ~ μ



Define n in terms of a perturbation about μ

$$n = \mu(1+\delta)$$

$$\delta << 1$$

Stirling's Approximation: $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ as $n \to \infty$

So, $$p(n\,|\,\mu) = \frac{\mu^n e^{-\mu}}{n!} \sim \frac{\mu^{\mu(1+\delta)} e^{-\mu}}{\sqrt{2\pi\mu(1+\delta)}\left(\frac{\mu(1+\delta)}{e}\right)^{\mu(1+\delta)}}$$

$$= \frac{\mu^{\mu(1+\delta)} e^{-\mu}}{\sqrt{2\pi\mu(1+\delta)}\left(\frac{\mu(1+\delta)}{e}\right)^{\mu(1+\delta)}} = \frac{\mu^{\mu(1+\delta)} e^{-\mu}}{\sqrt{2\pi\mu}\left[\mu^{\mu(1+\delta)}\right]\left[(1+\delta)^{\mu(1+\delta)+\frac{1}{2}}\right]\left[e^{-\mu(1+\delta)}\right]}$$

$$= \frac{e^{\mu\delta}}{\sqrt{2\pi\mu}(1+\delta)^{\mu(1+\delta)+\frac{1}{2}}} \equiv \frac{e^{\mu\delta}}{\sqrt{2\pi\mu}} \frac{1}{g}$$

Define: $f = \ln g = [\mu(1+\delta) + 1/2] \ln(1+\delta)$

Taylor Expand: $f' = \mu \ln(1+\delta) + [\mu(1+\delta) + 1/2]/(1+\delta)$

$(\delta \ll 1,\ \mu \gg 1)$

$$f'' = \frac{\mu}{1+\delta} + \frac{\mu}{1+\delta} - \frac{\mu(1+\delta)+1/2}{(1+\delta)^2}$$

$$f(0) = 0 \qquad f'(0) = \mu + \frac{1}{2} \simeq \mu \qquad f''(0) = \mu - \frac{1}{2} \simeq \mu$$

$$f \sim f(0) + f'(0)\delta + \frac{f''(0)}{2}\delta^2 = \mu\delta + \frac{\mu\delta^2}{2}$$

$$f \sim \mu\delta + \frac{\mu\delta^2}{2} \longrightarrow g \sim e^{\mu\delta + \mu\delta^2/2}$$

$$p(n\,|\,\mu) = \frac{e^{\mu\delta}}{\sqrt{2\pi\mu}}\frac{1}{g} \sim \frac{1}{\sqrt{2\pi\mu}}e^{\mu\delta - \mu\delta - \mu\delta^2/2}$$

$$= \frac{1}{\sqrt{2\pi\mu}}e^{-\mu\delta^2/2}$$

$$= \frac{1}{\sqrt{2\pi\mu}}e^{-(\mu\delta)^2/2\mu}$$

$$= \frac{1}{\sqrt{2\pi\mu}}e^{-(n-\mu)^2/2\mu}$$

recall:

$$n = \mu(1+\delta)$$
$$= \mu + \mu\delta$$

$$\mu \xrightarrow[\text{(Poisson)}]{} \sigma^2$$

$$\boxed{p(n\,|\,\mu) = \frac{1}{\sqrt{2\pi\sigma}}e^{-(x-\mu)^2/2\sigma^2}}$$

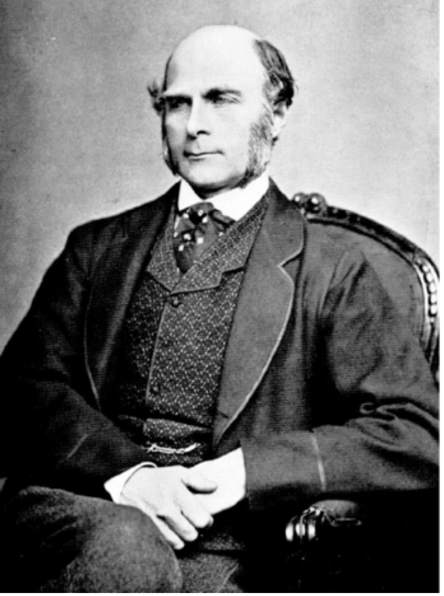
Central Limit Theorem

Pick N random numbers from an arbitrary distribution and define:

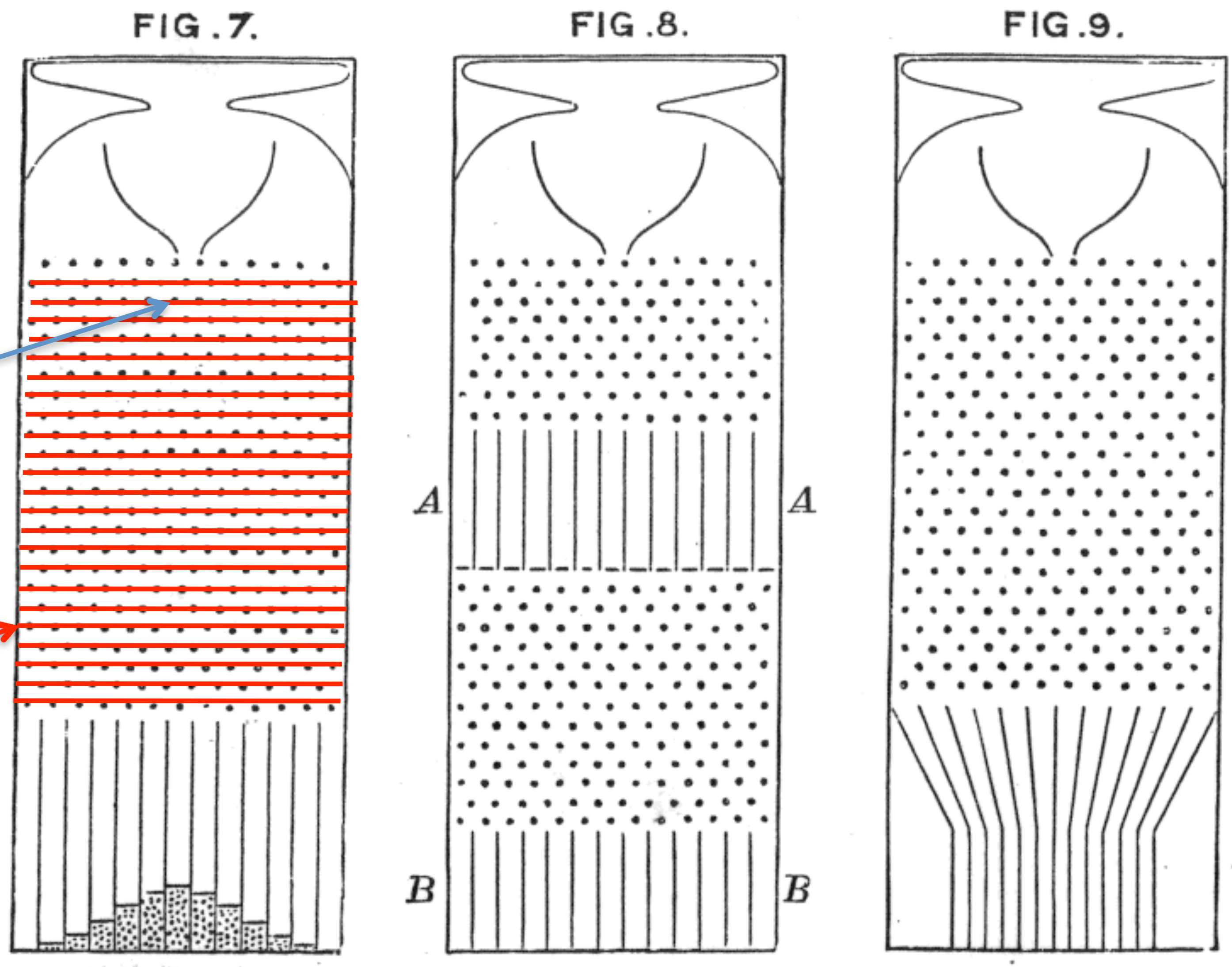
$$S = \sum_{i=1}^{N} x_i \quad \text{and} \quad \bar{X} = \frac{S}{N}$$

What is the probability distribution of $\boldsymbol{S}$ (or, equivalently, $\overline{X}$) ?

Galton Machine



Sir Francis Galton



ball has 50:50 chance of going right or left at each peg (underlying distribution)

"N" peg interactions (samples) get added to form the total deflection