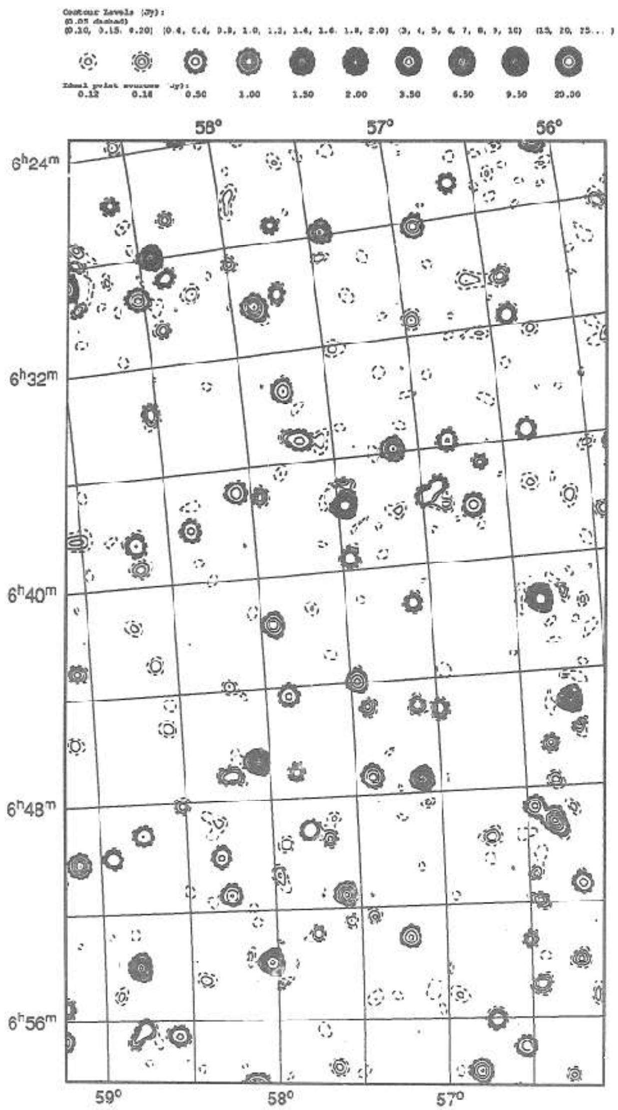
Lecture 6 – three useful probability distributions

- Binomial:
  - Identical trials with two possible outcomes, e.g. 1-d random walk, but many others.
- Poisson:
  - Radioactivity, shot-noise, mutation-rates, internet hits etc.
- Gaussian:
  - Johnson/thermal noise, many experimental errors.

Focus in this course is not on the maths but on their use & applications in physical systems and data analysis.

Physics is about inference – what do our measurements tell us about the universe?

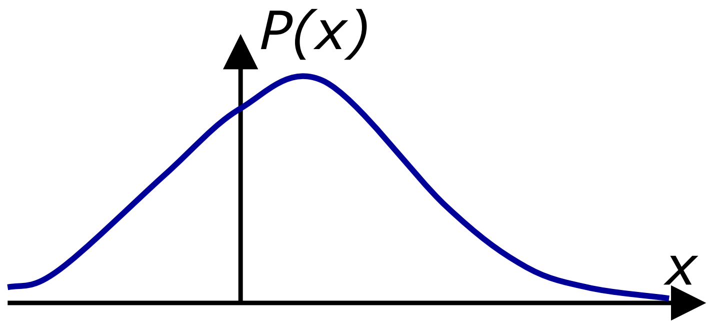


- Are these radio sources distributed randomly on the sky?

- To answer this question we might use probabilistic models to:
  - Model the phenomenon.
  - Characterize the noise.

Need to agree on what probability means: “degree of belief in the value of quantities/hypotheses”

- *P(x)dx* represents the probability that x lies in the range $x \rightarrow x+dx$.



- Clearly, x takes some value, so we normalize:

$$\int_{-\infty}^{+\infty} P(x)dx = 1.$$

- And we have the following expressions for average (expectation) values:

$$\langle x \rangle = \int_{-\infty}^{+\infty} x\, P(x)\, dx\,,$$

$$\langle x^2 \rangle = \int_{-\infty}^{+\infty} x^2 P(x)\, dx\,, \text{etc.}$$

- If we are to associate probability with our degree of belief, then:
  - How do we assign these probabilities?

How we establish our degree of belief can rely upon many approaches:

- The "Principle of indifference" – we assign equal probabilities to events unless we have information distinguishing them.
  - This approach derives from Laplace.

- We can assign probabilities on the basis of the relative frequencies of events in the limit of a large number of trials:
  - This is often called the "frequentist" approach.

- We simply examine what information we have and use that:
  - E.g. there have been 5 naked eye supernovae in the past 10 centuries: let's use this to assign a quantitative measure of our belief that another supernova will occur in the next century.
  - This is often called the "Bayesian" approach.
  - NB – all observers with the same information must agree on their probabilities.

The Binomial distribution is associated with identical trials with only 2 possible outcomes

- Call the probability of one outcome ("success") *p*, and the other, "failure", *(1-p)*, e.g tossing a fair coin:
  - p_{suc}= ½ for heads, and p_{fail} = 1-*p* = ½ for tails.

- Consider *N* independent trials, with *r* "successes" & (*N*-*r*) "failures":
  - E.g the sequence "s s s s s…s f f f f f f".
  - The probability of this arrangement is clearly $p^r (1-p)^{N-r}$.
  - But we can permute this arrangement and still retain the same numbers of successes and failures.
  - The number of permutations is $= \dfrac{N!}{r!(N-r)!} \equiv {}^N C_r \equiv \dbinom{N}{r}$.

- So, the probability of getting r successes from N trials is given by:

$$\boxed{\mathrm{Prob}(r|p,N) = p^r (1-p)^{N-r} \frac{N!}{r!(N-r)!}.}$$

Let’s check the normalization, mean and variance of this result

- The distribution is already normalized, since:

$$\sum_{r=0}^{r=N} p^r (1-p)^{N-r} \frac{N!}{r!(N-r)!} = \left[p+(1-p)\right]^N = 1.$$

- The mean number of successes <*r*> is given by:

$$\langle \mathrm{r} \rangle = \sum_{r=0}^{r=N} r\, \mathrm{Prob}(r|p,N) = \sum_{r=\cancel{0}1}^{r=N} r\; p^r (1-p)^{N-r} \frac{N!}{r!(N-r)!}.$$

r=0 term is zero, pull out *Np*.

$$= Np \sum_{r=1}^{r=N} p^{r-1} (1-p)^{N-r} \frac{(N-1)!}{(r-1)!(N-r)!}.$$

Finally, substitute r’=r-1 and N’=N-1 to get:

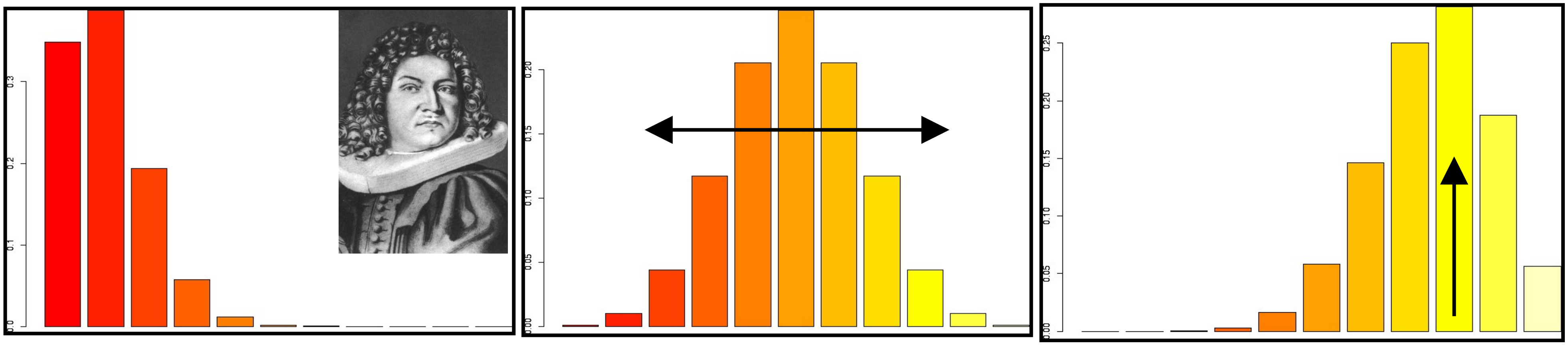
$$\boxed{\langle r \rangle} = Np \sum_{r'=0}^{r'=N'} p^{r'} (1-p)^{N'-r'} \frac{N'!}{r'!(N'-r')!} = Np \times \left[p+(1-p)\right]^{N'} = \boxed{Np\,.}$$

- We can compute the variance of r in the same way, getting:

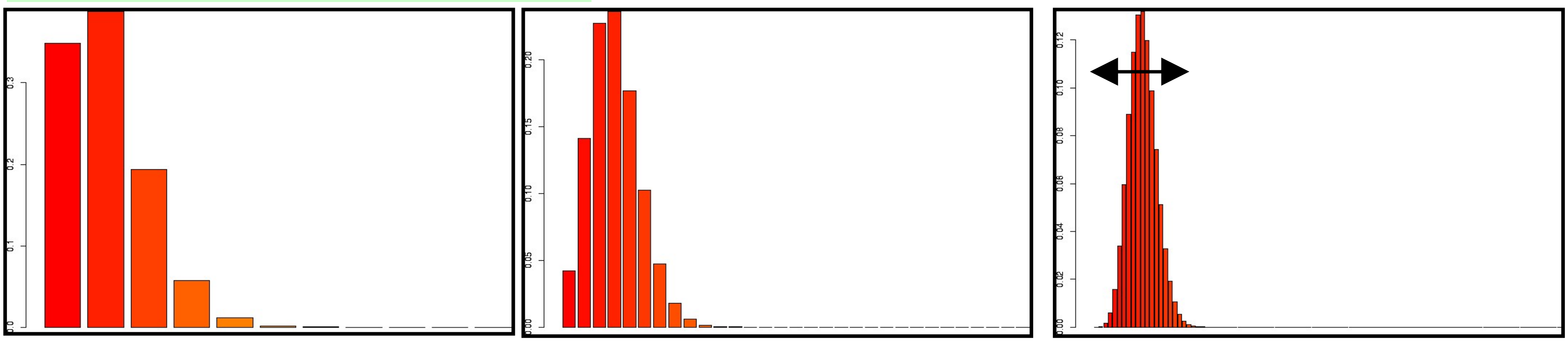
$$\boxed{\mathrm{Var}(r) = \langle r^2 \rangle - \langle r \rangle^2 = Np\,(1-p)\,.}$$

What is the behaviour as $p_{success}$ and N_{trials} change?

N=10, p = 0.1, 0.5, 0.75



p=0.l, N=10, 30, 100



What you need to recall about Binomial processes in physics

- For r "successes", each of probability p, and $(N-r)$ "failures", each with probability $(1-p)$;
  - $\langle r \rangle = Np$.
  - Standard deviation of $r = [Np(1-p)]^{1/2}$
- Binomial distributions:
  - Peak *<u>around</u>* the mean $\langle r \rangle = Np$.
  - As N increases, the peak width slowly narrows with respect to the range of N.
  - The peak width also depend on p – it's widest when p=0.5.
  - Are skew, when $p \neq 0.5$.
- Use when $\exists$ N trials, each of which has 2 outcomes, e.g.:
  - 1-d random walk;
  - System of non-interacting spins in a lattice.

Sometimes we do not “run” the trials

- We know the number of outcomes, but not the number of trials – e.g. we hear the clicks of a Geiger counter but <u>not</u> its “non-clicks”:
  - Other examples include lightning flashes, raindrops on a roof, discrete charges passing, photons arriving from a lamp.
  - These events are definite, but happen only sometimes, and at random.
  - If we measure how many occur over a short time, it fluctuates.

- Such processes are described by the <u>Poisson</u> distribution:
  - The key idea is that the phenomenon is characterised by a <u>rate</u>, i.e. the rate that would be determined if we waited for a <u>long</u> time and measured the number of events over that time. NB – we assume the mean rate is constant.

- E.g.: if we measured 500 photons in one second, would we be surprised if we measured 400 (or 600) in the next?

The Poisson dist. is the Binomial dist. in the limit of rare events, and many trials

- Consider the Binomial distribution when $p \to 0$ and $N \to \infty$, such that $\langle r \rangle = Np$ stays finite.

- Imagine we have λ events per unit interval. Let's split the interval into N tiny equal sections such that the chance of getting 2 events per section $\to 0$.
  - The probability that any section contains an event is $p=\lambda/N$.
  - Then, application of our results for the Binomial distribution for the total number of events over the interval gives:

$$\mathrm{Prob}(r|p,N) = p^r (1-p)^{N-r} \frac{N!}{r!(N-r)!} \to$$

$$\boxed{\mathrm{Prob}(r|\lambda/N, N) = \frac{\lambda^r}{N^r}\left(1-\frac{\lambda}{N}\right)^{N-r} \frac{N!}{r!(N-r)!}.}$$

Now let $N \rightarrow \infty$ with r finite:

$$\text{Prob}(r|\lambda/N, N) = \frac{\lambda^r}{N^r}\left(1-\frac{\lambda}{N}\right)^{N-r}\frac{N!}{r!(N-r)!}$$

- Find that: $$\boxed{\text{Prob}(r\,|\,\lambda) = \frac{\lambda^r}{r!}e^{-\lambda}}$$ for the probability of r events given a mean event number λ.

- Here we have used:

$$\frac{N!}{r!(N-r)!} = \frac{N(N-1)(N-2)\ldots(N-[r-1])}{r!} \rightarrow \frac{N^r}{r!} \quad \text{and}$$

$$\left(1-\frac{\lambda}{N}\right)^{N-r} \rightarrow \left(1-\frac{\lambda}{N}\right)^{N} \rightarrow e^{-\lambda} \text{ (defn. of } e^x\text{) .}$$

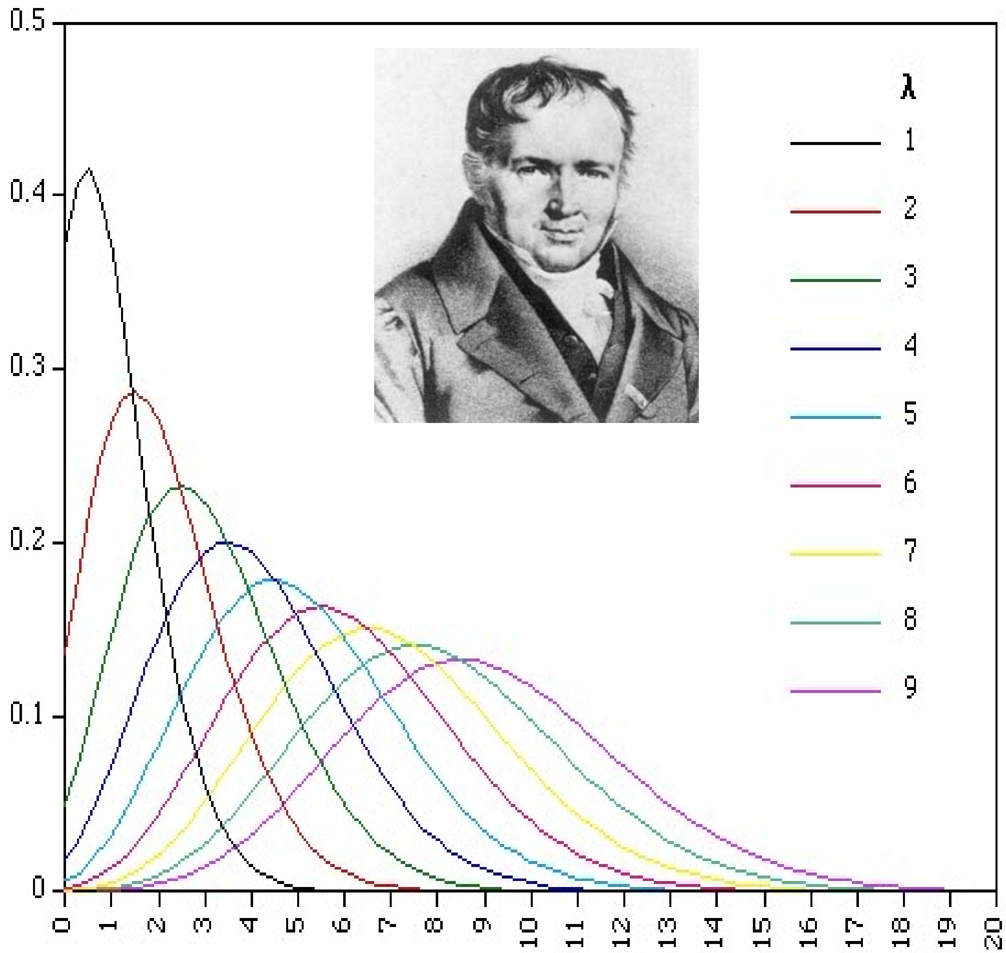
- Normalization: $$\sum_{r=0}^{r=\infty}\text{Prob}(r|\lambda) = e^{-\lambda}\sum_{r=0}^{r=\infty}\frac{\lambda^r}{r!} = e^{-\lambda}e^{\lambda} = 1.$$

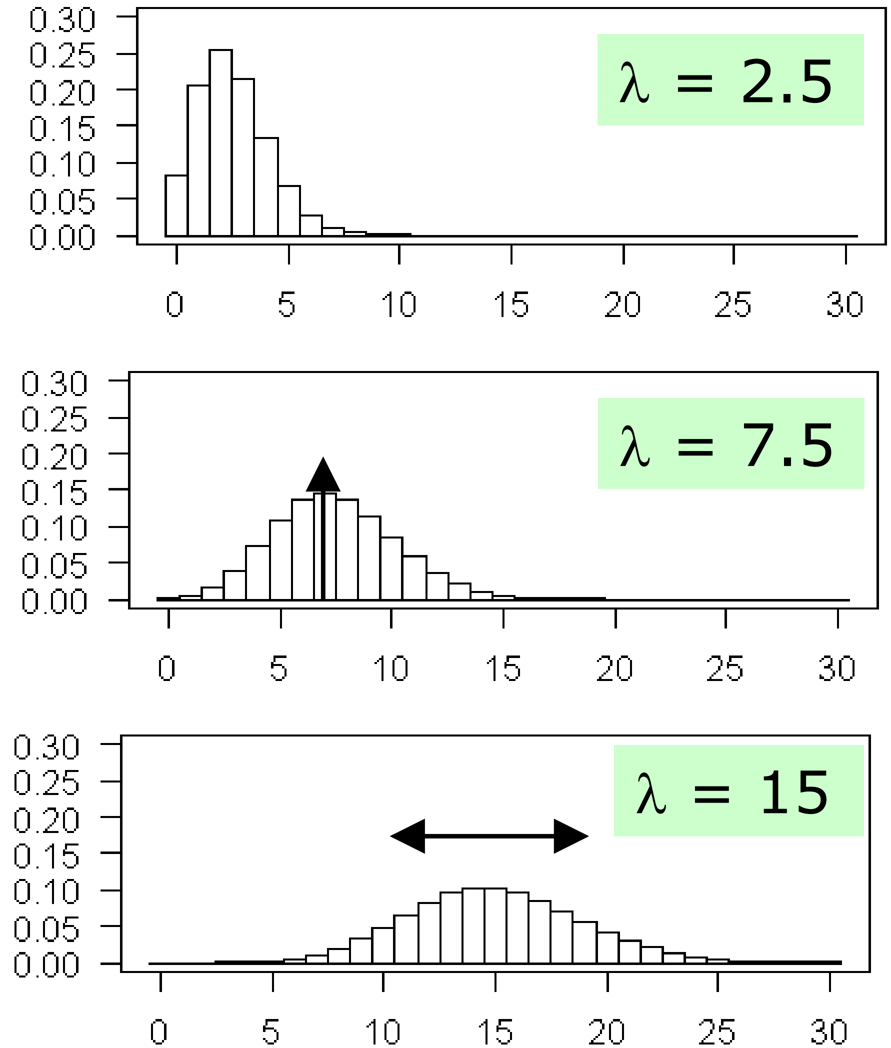
- The mean of r: $$\langle \text{r} \rangle = \sum_{r=0}^{r=\infty} r\frac{\lambda^r}{r!}e^{-\lambda} = \lambda.$$

- Variance of r: $$\text{Var}(r) = \langle r^2 \rangle - \langle r \rangle^2 = \lambda.$$

Note these are the same

What happens as λ (the mean rate) changes?





What you need to recall about Poisson processes in physics

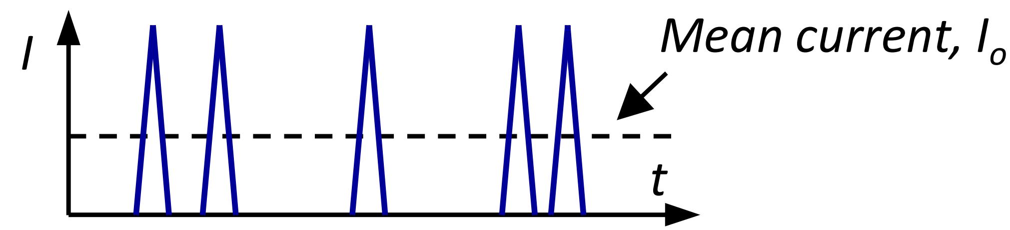
- Events occur randomly, but with a known mean rate, λ, over some interval (in time or space), and independent of when the last event occurred.

- The mean number, $<r>$, of events measured in the interval is λ, which equals the variance in the number of events.
  - $<r> = \lambda$.
  - Standard deviation of $r = \lambda^{1/2}$.

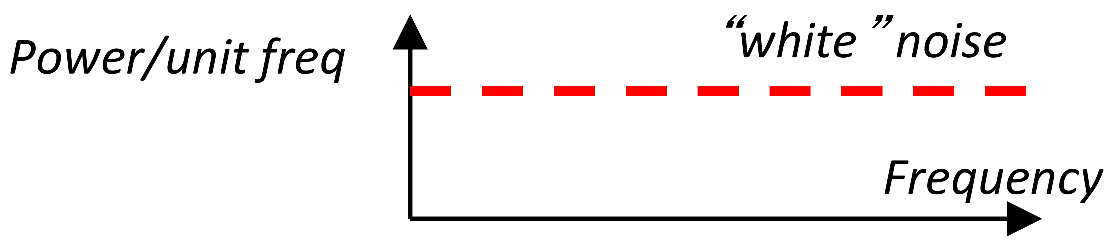
- In comparison with the Binomial, where the variance, *np(1-p),* is always less than the mean, *np*:
  - Poisson is broader.
  - Poisson has a characteristic long upper tail.
  - The number of events actually observed can greatly exceed the mean event rate, whereas for the Binomial, the number of successes, *r*, cannot exceed the number of trials, *N*.

An example Poisson process: the level of fluctuations in a current signal – "shot noise"

- Consider a current composed of discrete charges arriving at random times:



  - Given that each arrival pulse is very short, the frequency spectrum of I(t) must include very high frequencies. In fact we have:



- A single electron arriving in $\Delta t \Rightarrow$ instantaneous current $e/\Delta t$.
- So if N electrons arrive at random in Δt *(on average)* there will be an associated net current fluctuation $\Delta I \approx \sqrt{N} \times (e/\Delta t)$.

Shot noise - continued

$$\Delta I \approx \sqrt{N} \times (e/\Delta t)$$

- So, remembering that the average current is $I_{Average} \approx \dfrac{Ne}{\Delta t}$,

  we find: $$\overline{\Delta I^2} \approx \frac{Ne^2}{[\Delta t]^2} = \frac{I_{Average}\, e}{\Delta t}.$$

- Now, a range of frequencies, $\Delta\nu = 1/\Delta t$ will be present, so we can write: $$\overline{\Delta I^2} \approx I_{Average}\, e\, \Delta\nu.$$

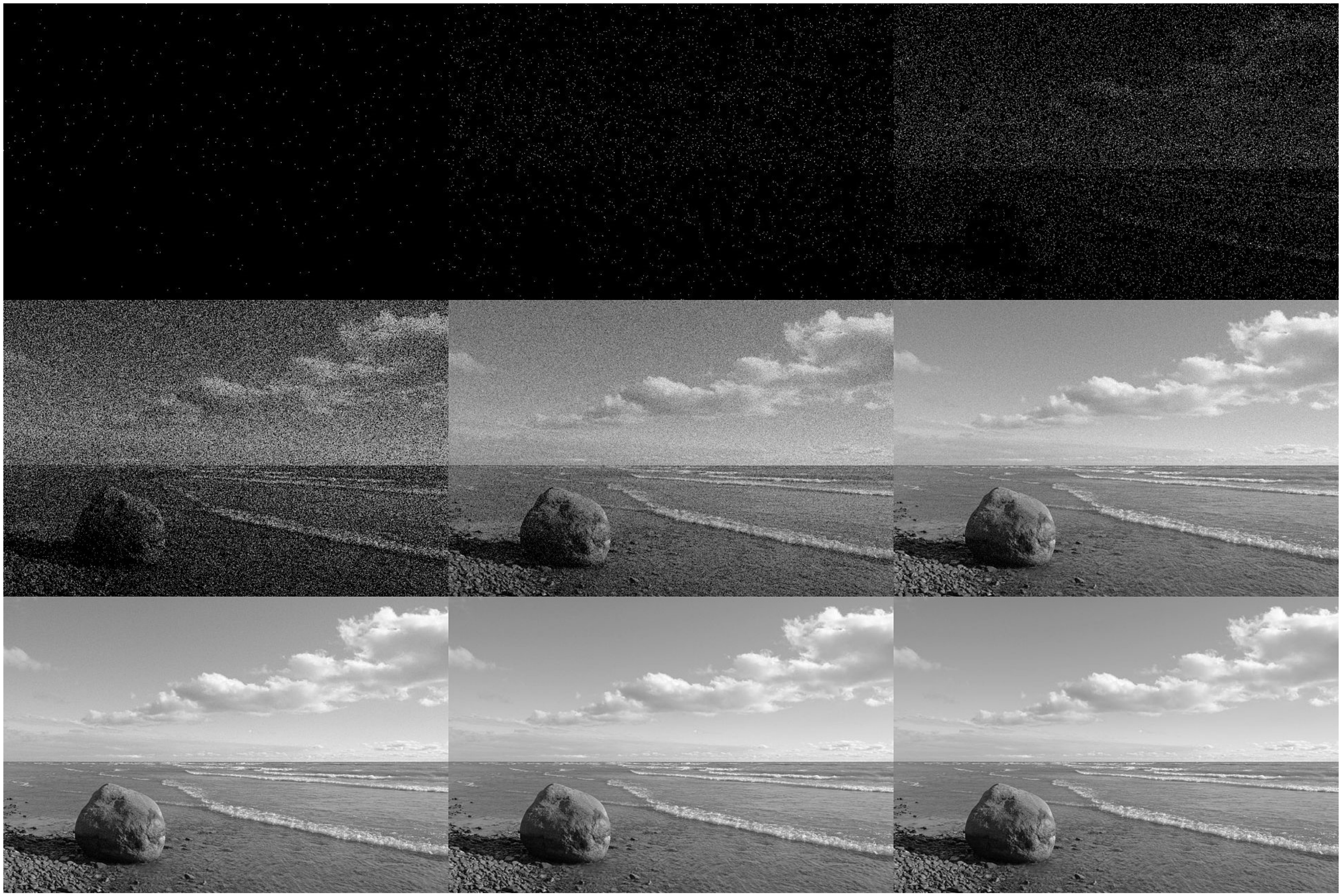
- A rigorous treatment gives: $$\boxed{\overline{\Delta I^2} \approx 2 I_{Average}\, e\, \Delta\nu.}$$

- So rms. "shot noise" current $\Delta I_{rms} = [2\, I_{average}.\, e\, .\text{bandwidth}]^{1/2}$.

  For $I_{Average}$ = 1 A and B=10 kHz: (10[14] e[-] in 100 μs) $$\frac{(\Delta I)_{rms}}{I_{Average}} = 0.000006\ \%.$$

  For $I_{Average}$ = 1 pA and B=10 kHz: (600 e[-] in 100 μs) $$\frac{(\Delta I)_{rms}}{I_{Average}} = 6\%.$$

Photon noise – rate increasing from top left to bottom right



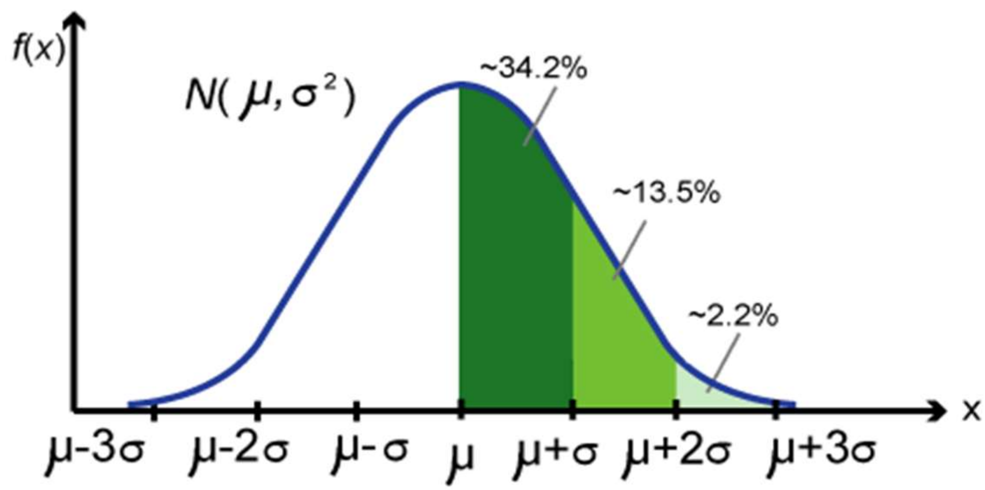
The Gaussian distribution

$$\text{Prob}(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}.$$

Mean

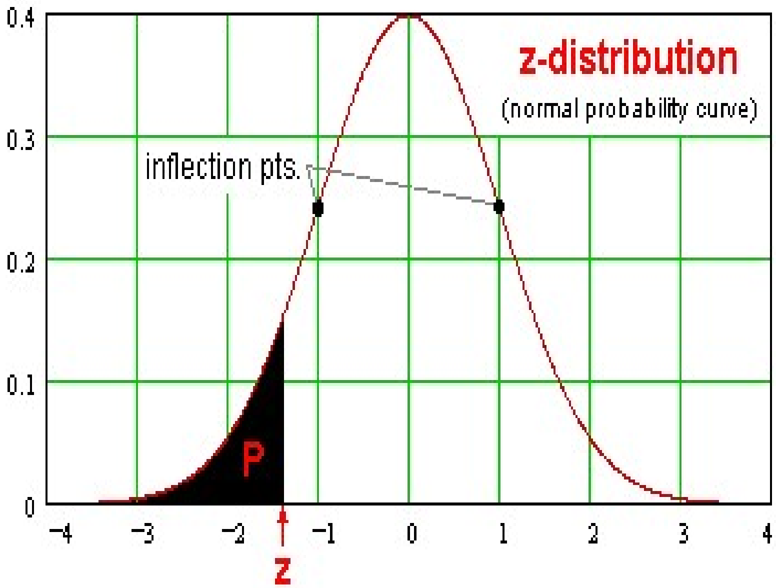
Standard deviation

- Sometimes referred to as Normal or “bell” curve.



Useful data

$$Here\ , z = \frac{(x-\mu)}{\sigma}.$$



	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Properties of the Gaussian distribution

- Apart from the shift by μ and scaling by σ, the shape is always the same, hence the substitution $z = (x-\mu)/\sigma$.
- This gives the unit Gaussian: $$P(z)dz = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz .$$
- It is relatively easy to show that:
  - Prob($x|\mu,\sigma$) is indeed normalized.
  - Mean of x $$= \int_{-\infty}^{+\infty} x \,\mathrm{Prob}(x|\mu,\sigma)\, dx = \mu.$$
  - Variance of x $$= \int_{-\infty}^{+\infty} x^2 \,\mathrm{Prob}(x|\mu,\sigma)\, dx = \sigma^2.$$

Use $\int_{-\infty}^{+\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{\mathrm{a}}}$, integration by parts multiple times etc.

The Gaussian is of immense use in describing distributions of experimental measurements, and much else. Why?

First: the Gaussian can be considered as the limit of a Poisson process with λ large:

- Consider a Poisson distribution, with $P(r|\lambda) = e^{-\lambda}\lambda^r/r!$
  - If we take the ln of each side, then using Stirling's approximation, we obtain:

$$\boxed{\ln r! \approx -r + r\ln r + \ln\sqrt{2\pi r}}$$

$$\ln\left[\mathrm{Prob}(r|\lambda)\right] \approx -\lambda + r\ln\lambda - \left(-r + r\ln r + \ln\sqrt{2\pi r}\right).$$

  - Now put $\varepsilon = r - \lambda$, since we expect $P(r|\lambda)$ to peak around $r=\lambda$:

$$LHS \approx -\lambda + (\lambda+\varepsilon)\left\{\ln\lambda - \ln\left[\lambda\left(1+\frac{\varepsilon}{\lambda}\right)\right]\right\} + (\lambda+\varepsilon) - \ln\sqrt{2\pi(\lambda+\cancel{\varepsilon})}.$$

$\varepsilon \ll \lambda$

  - Using the expansion $\boxed{\ln(1+z) \approx z - z^2/2 + \ldots}$

$$LHS \approx \varepsilon - (\lambda + \cancel{\varepsilon})\left(\frac{\varepsilon}{\lambda} - \frac{\varepsilon^2}{2\lambda^2}\right) - \ln\sqrt{2\pi\lambda}$$

$$\approx -\frac{\varepsilon^2}{2\lambda} - \ln\sqrt{2\pi\lambda},$$

where we have again assumed $\varepsilon \ll \lambda$.

The Gaussian limit of a Poisson distribution, cont[d.]

$$\ln(\mathrm{Prob}\ (r \mid \lambda)) \approx -\frac{\varepsilon^2}{2\lambda} - \ln\sqrt{2\pi\lambda},$$

- Exponentiating both sides, we finally we arrive at:

$$\mathrm{Prob}(r \mid \lambda) = \frac{e^{-(r-\lambda)^2/2\lambda}}{\sqrt{2\pi\lambda}}, \text{i.e. a Gaussian with } \mu = \lambda \text{ and } \sigma = \sqrt{\lambda}.$$

- So, a Gaussian is a good approximation to a Poisson distribution, provided λ is large enough – say >10.

- In fact, the Binomial distribution also tends to a Gaussian as *N* gets large.

- This is the first reason why the Gaussian is so useful.

The Central Limit Theorem

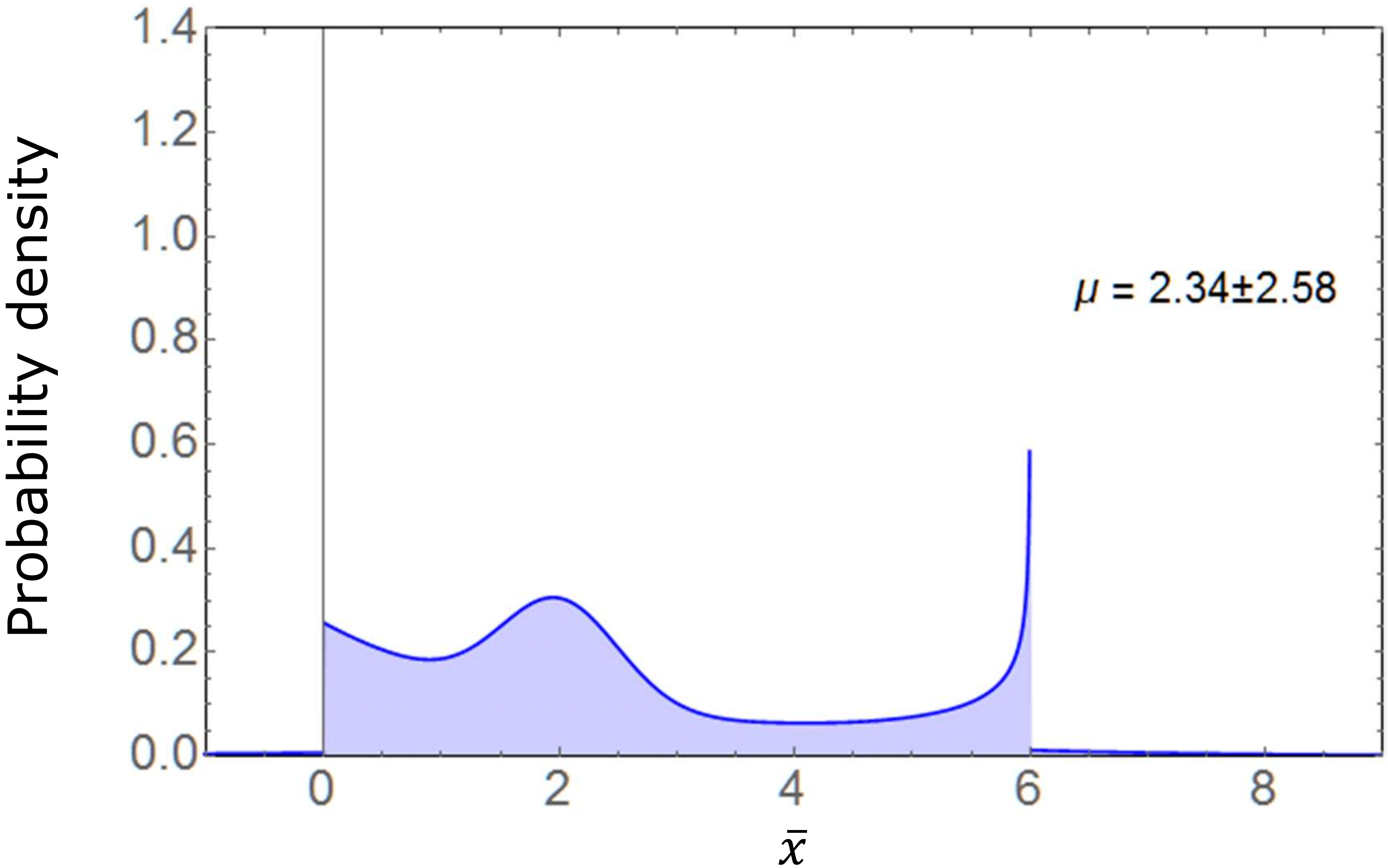
- Suppose we have N samples of some different (independent) quantities, each indicated by a subscript i, each drawn from a different $P_i(x)$.
- Let each sample distribution have mean μ_i and variance σ_i^2.
- Then, if we add the samples:
  - The resulting sample mean is = $\Sigma\,\mu_i$.
  - The resulting sample variance is = $\Sigma\sigma_i^2$.
  - The probability distribution describing the sum becomes Gaussian as $N \rightarrow \infty$.

A quantity that depends on the sum of many independent variables will tend to have a Gaussian probability distribution.

- So, multiple causes of error, tend to lead to a measured quantity having a Gaussian error distribution.

[Analytical proof is hard, but numerical illustration is easy.]

Central Limit Theory in action

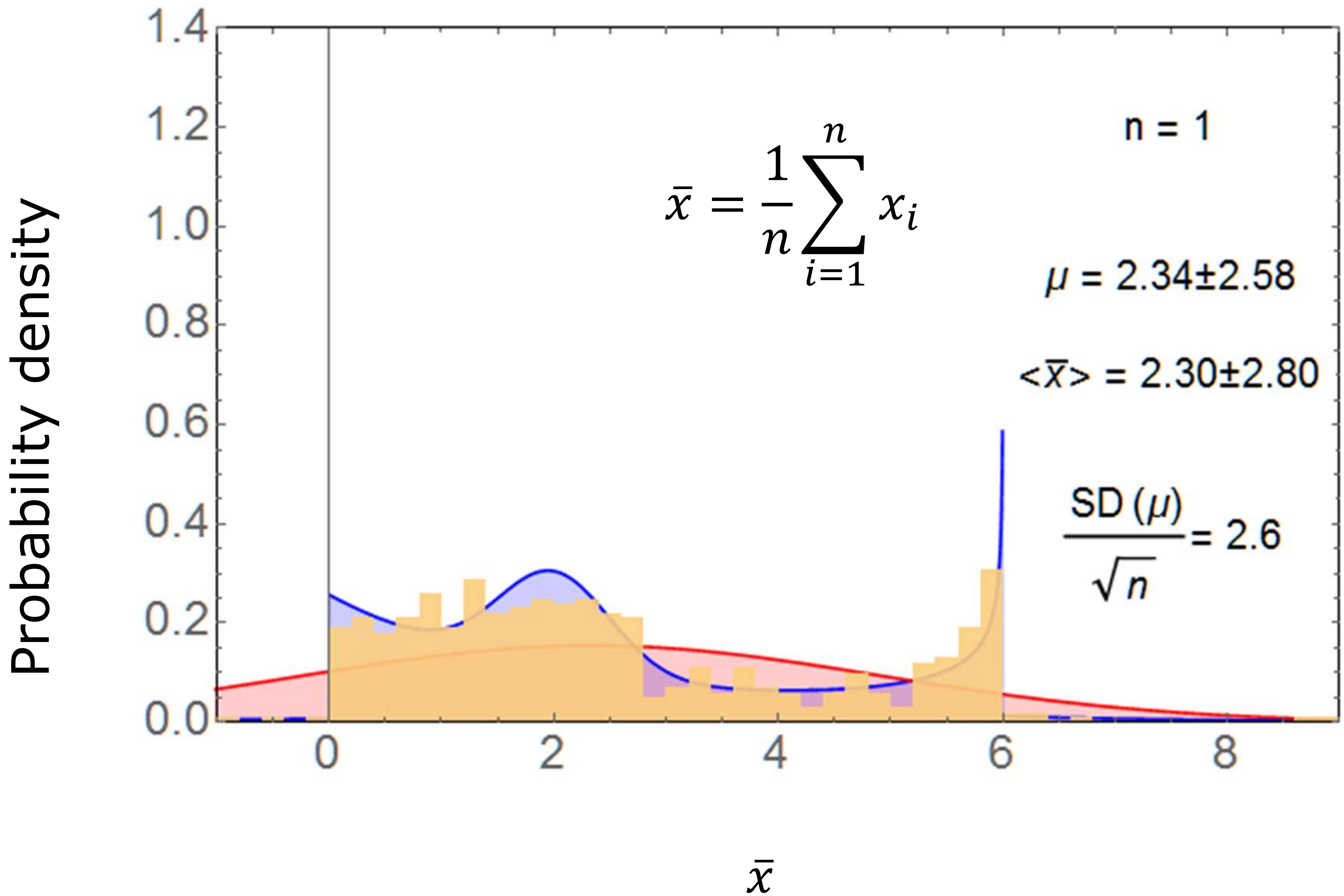


Draw a number from the distribution n times and taken an average ($\bar{x}$)

Each pick must be independent of previous draws.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The beige cells show the distribution of $\bar{x}$



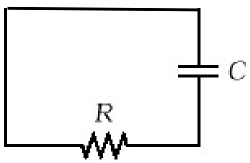
A third reason is that a number of important physical processes exhibit Gaussian statistics

- "Johnson" or "thermal" noise: associated with the agitation of a system in thermal equilibrium at some temp T.
- Occurs when any term in the expression for the system energy includes a coordinate squared. Each quadratic term has an average energy = ½kT. (equipartition theory)

- E.g., a crystal (or resistor, or piece of wire) at some given T will exhibit lattice vibrations, and these will move the lattice electrons back and forth to give an rms current:

$$\frac{1}{2}kT = \left\langle \begin{matrix}\text{Energy in atom} \\ \text{vibration}\end{matrix} \right\rangle = \left\langle \frac{1}{2} m v_{x,\,atom}^{2} \right\rangle.$$

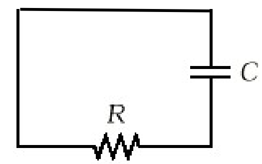
- This can manifest itself in different ways. In this circuit, you see a fluctuating voltage across the capacitor:



Although we can use a circuit to help analyze the situation, it is T (temp) that matters

- In the case of our RC circuit, we can write:

$$\frac{1}{2}kT = \left\langle \frac{1}{2}CV^2 \right\rangle \Rightarrow \overline{V^2} = \frac{kT}{C}.$$



- The time constant, τ, = RC, so the frequencies of relevance correspond to a bandwidth, $\Delta\nu$, = $1/\tau$ = 1/RC.
- Hence $\overline{V^2} = \frac{kT}{C} = kT\,R\,\Delta\nu$ and so the rms thermal noise power $\overline{V^2}/R = kT\,\Delta\nu$, independent of R.
- A rigorous treatment gives $\boxed{\text{Power} = \overline{V^2}/R = 4kT\,\Delta\nu.}$
  - E.g, if bandwidth $\Delta\nu$ = 10 kHz, and T = 300 K:
  - rms thermal power is $4 \times 1.6\times10^{-23} \times 300 \times 10^4 \approx 2 \times 10^{-16}$ W.
  - So, for I_o=1 pA, rms noise voltage ΔV_{rms} ~ few hundred μV.

Summary so far

- Binomial: N trials with 2 possible outcomes, probs *p* and *1-p:*
  - Important physical examples include random walk, spins.
- Poisson: only know mean event rate, λ:
  - Has mean = variance = λ
  - Important physical examples include shot noise.
- Gaussian: characterized by mean and sigma:
  - Is limiting case for Binomial for large N.
  - In limiting Poisson case (λ large), sigma is $\sqrt{\lambda}$.
  - CLT => frequently associated with errors.
  - Important physical examples include Johnson noise, other randomized fluctuations.

In our final lecture we apply these ideas to parameter estimation, goodness of fit and hypothesis testing.