Math Methods and Experimental Physics

Distributions

Distributions

There are tons and tons of interesting, useful distributions. I will only describe a handful:

- Binomial
- Poisson
- Normal
- Uniform

Average, variance, other moments

You have a set of N measurement x_i. You can calculate the average:

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

And you can calculate the variance:

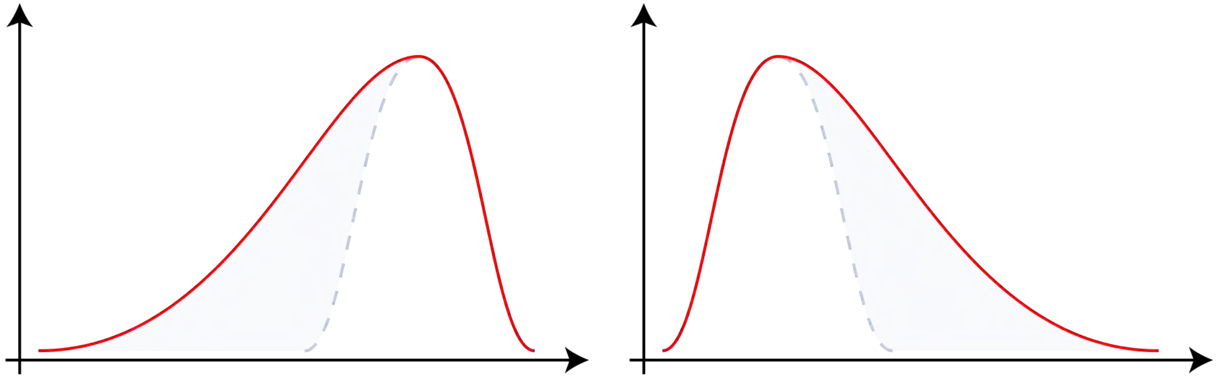
$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \rightarrow \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

And normalized moments:

$$E[X^n] = \frac{1}{N\sigma^n}\sum_{i=1}^{N}(x_i - \bar{x})^n$$ (biased)

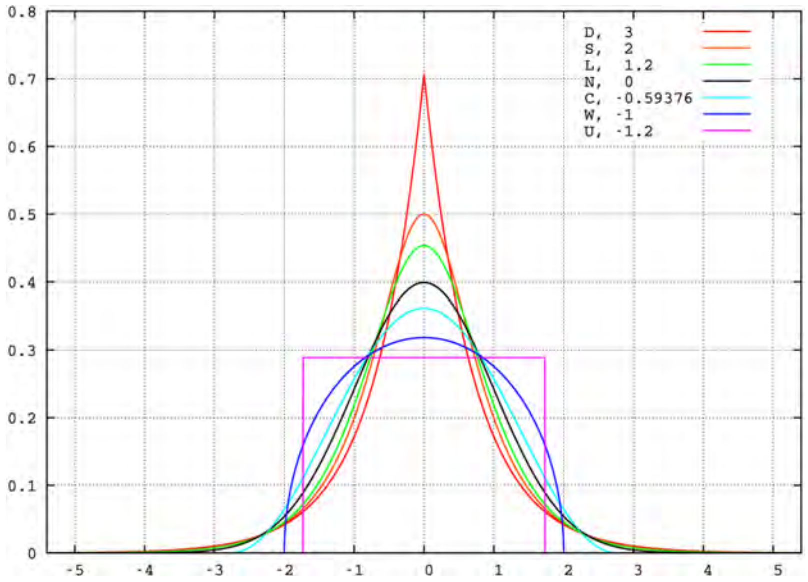
Third moment: skewness. Fourth moment: kurtosis

Average, variance, other moments



Negative Skew

Positive Skew



Note that N, r are integer
The binomial distribution is discrete

Binomial Distribution

Imagine an experiment with 2 outcomes, called “success” and “failure”. Success has a per-experiment probability p and failure 1-p.

Out of a fixed number, N, trials/experiments, what is the probability of r successes?

$$P_N(r) = \left(\begin{array}{c} r \\ N \end{array}\right) p^r (1-p)^{N-r} = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r}$$

Ways to combine r successes and N-r failures

r successes

N-r failures

Binomial Distribution

Assuming we know p, we can calculate the expectation of r and the variance:

$$\bar{r} = \sum_{r=0}^{N} r P(r) = Np$$

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

But we can know that the process is binomial, while p is unknown:

$$\bar{p} = \frac{\bar{r}}{N}$$

(See that this only makes sense if we have a way to estimate $\bar{r}$, e.g. a frequentist approach)

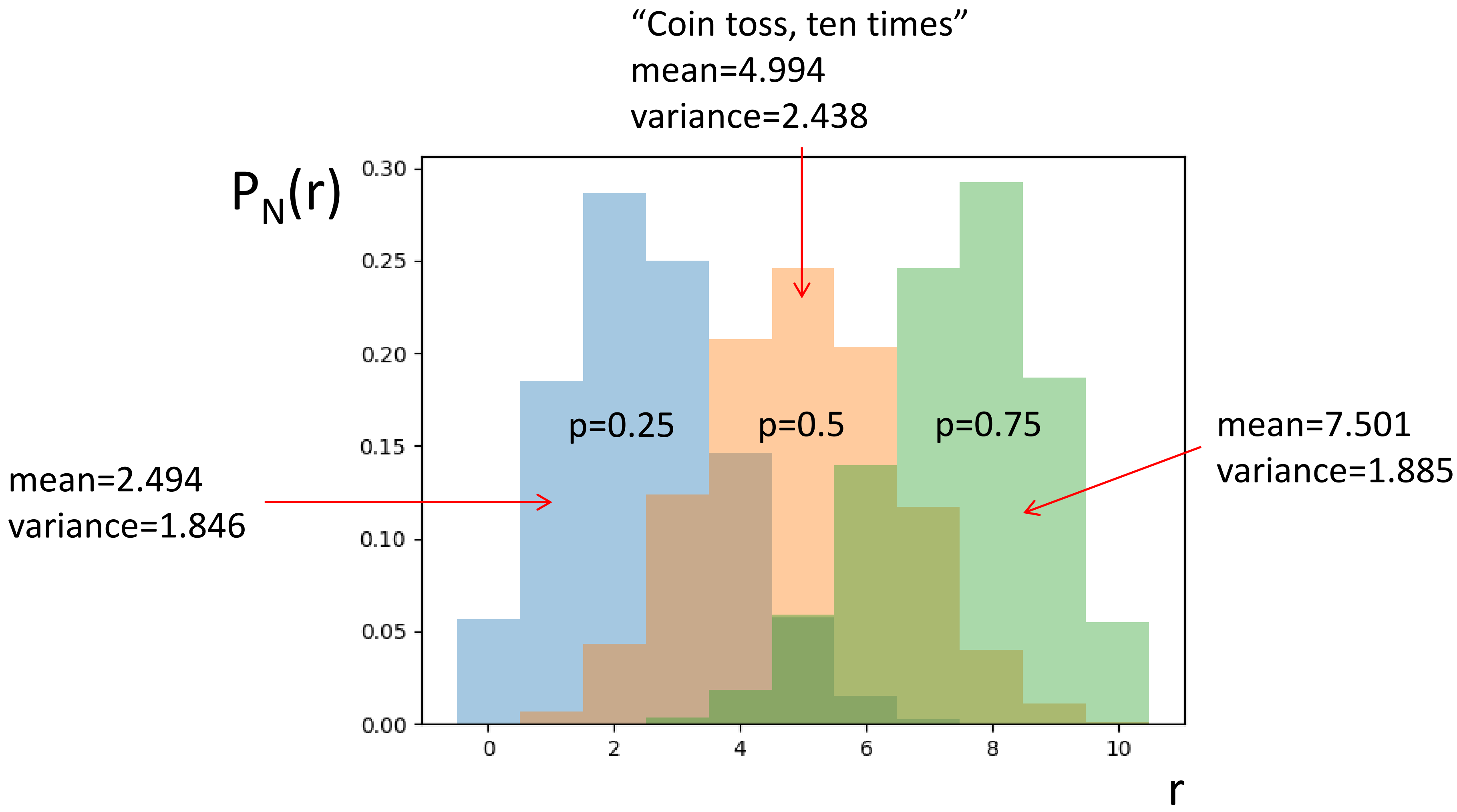
Where s^2 is a point estimate of the variance.

$$s^2 = \frac{N}{N-1}\bar{r}(1 - \frac{\bar{r}}{N})$$

Unbias term

Binomial Distribution

N = 10
p = 0.25, 0.5, 0.75
10,000 trials



Do I have to work on a weekend?

IceCube reports, real time, neutrinos that are likely to be astrophysical. We review these alerts after they are sent. From April 2016 to October 2018 IceCube sent 16 alerts, of which 11 fell on a weekend. What's the probability of that?

$$p_{weekday} = 2/7 \qquad p_{weekend} = 5/7$$

$$P_{16}(11) = \frac{16!}{11!5!}(2/7)^{11}(5/7)^5 = 8.4 \times 10^{-4}$$

A related question: how probable is that we got *at least*, that unlucky?

$$P_{16}(11) + P_{16}(12) + P_{16}(13) + P_{16}(14) + P_{16}(15) + P_{16}(16) = 10^{-3}$$

This last quantity is called a p-value.

Poisson Distribution

Note that n is integer
The Poisson distribution is discrete

An event can occur 0,1,2,3, … many times* in a very large set of trials. The average number of occurrences is μ. Then the probability of observing n, given μ is:

$$P_\mu(n) = \frac{\mu^n}{n!} e^{-\mu}$$

"Count" experiments are frequently described by Poisson statistics. Given μ, the expected measurement is: $\bar{n} = \mu$

And the variance is: $\sigma^2 = \mu$

This gives rise to the counting "statistical" uncertainty: $N \pm \sqrt{N}$

If μ is not known, then $\bar{n}$ serves as expectation and mean and variance point estimates.

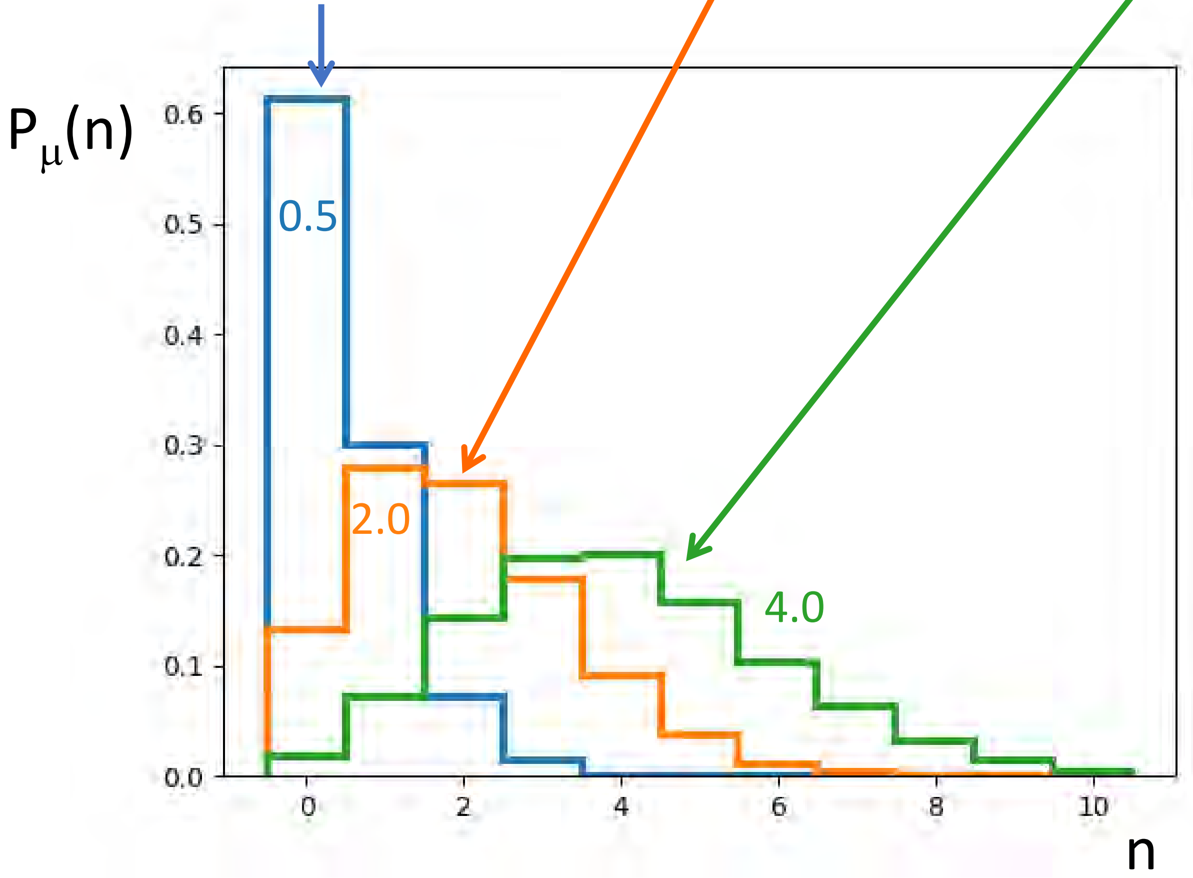
(*) If there's an upper bound, N, a binomial distribution may be more appropriate.

Poisson Distribution

Point Estimates → mean=0.493 variance=0.499

mean=2.002 variance=2.017

mean=4.021 variance=3.924



Histograms: python (numpy) generated. 1000 samples.

Colliding with asteroids

On average, a spaceship collides with one asteroid in a 100-year mission. What is the probability of not-being hit over 10 years?

$$\mu = 0.1 = R\Delta T$$

$$P_{\mu=0.1}(0) = 0.904$$

Which means that the probability of being hit is:

$$P_{\mu=0.1}(\geq 1) = 0.095$$

<u>Discuss</u>: why is $P_\mu(\geq 1)$ different than 0.1 ?

Colliding with asteroids

On average, a spaceship collides with one asteroid for a 100-year mission. What is the probability of not-being hit over 10 years?

$$\mu = 0.1 = R\Delta T$$

$$P_{\mu=0.1}(0) = 0.904$$

Which means that the probability of being hit is:

$$P_{\mu=0.1}(\geq 1) = 0.095$$

Discuss: why is $P_\mu(\geq 1)$ different than 0.1 ? For small μ:

$$P_\mu(0) = e^{-\mu} \qquad P_\mu(\geq 1) = 1 - e^{-\mu} \sim \mu$$

Normal (Gaussian) distribution

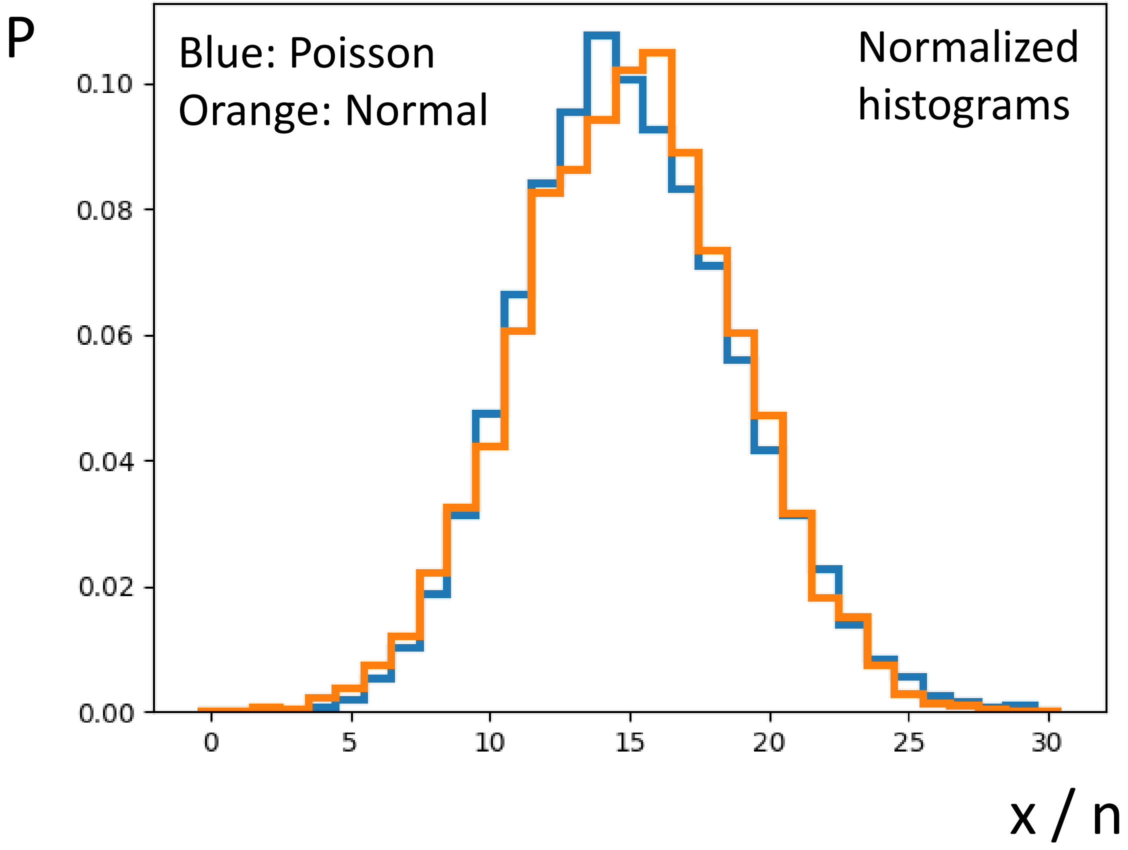
Note that x is real
The normal distribution is <u>continuous</u>: probability density function

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

Here x is a random real-valued variable, μ is the mean and σ is the standard deviation (the variance is σ^2)

A Gaussian probability of mean μ and variance μ, for integer-valued random variables, is the limit of $\mu \to \infty$ of the Poisson distribution.

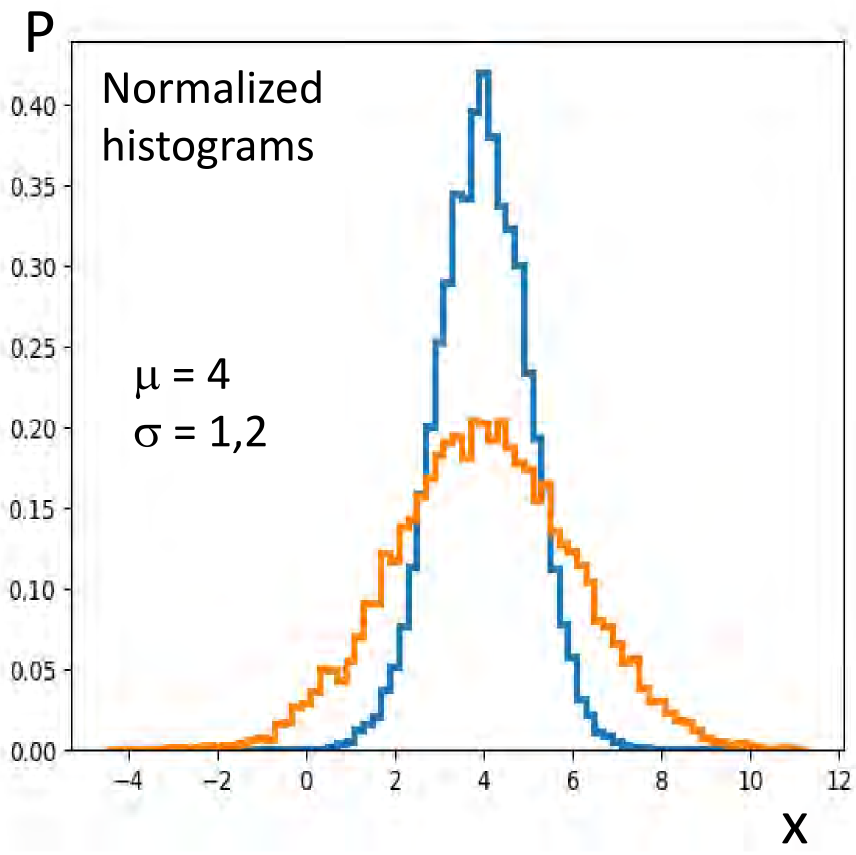
Normal vs. Poisson

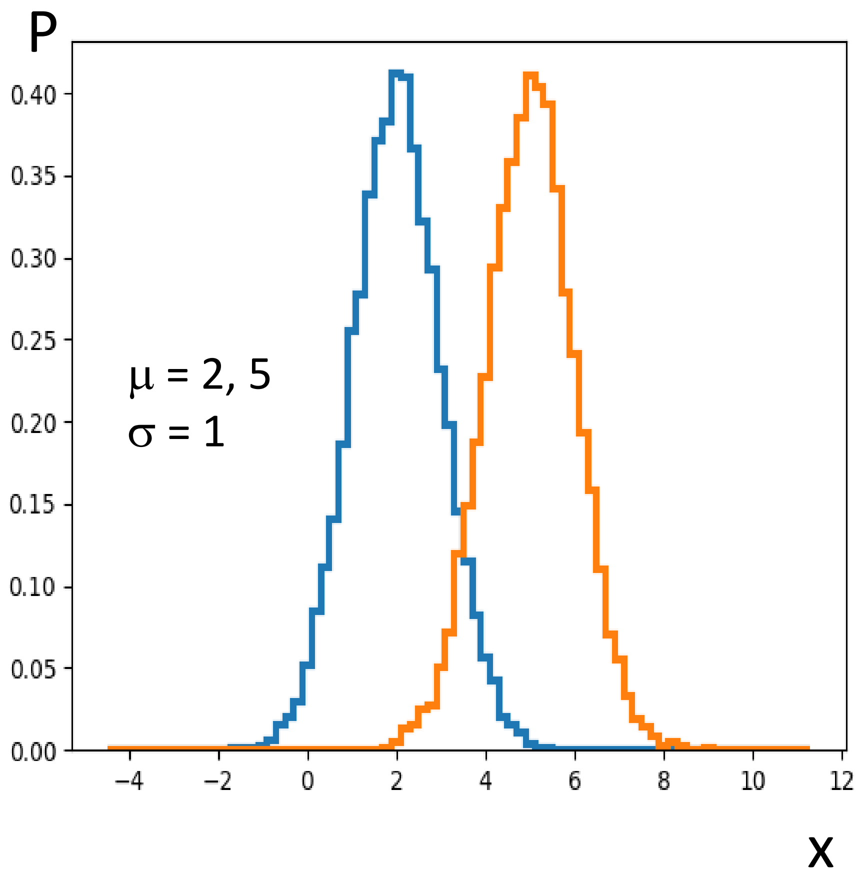


Poisson:
μ = 15

Normal:
μ = 15
σ^2= 15
10,000 trials

Normal distribution





Error function

The error function is: $erf(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt$

Cumulative Distribution Function for the normal distribution

The normal cumulative distribution function is:

$$\frac{1}{2}\left[1 + erf(\frac{x-\mu}{\sqrt{2}\sigma})\right]$$

Integrating from the left or -∞ to x.

… and if you integrate from the right:

$$\frac{1}{2}\left[1 - erf(\frac{x-\mu}{\sqrt{2}\sigma})\right]$$

Integrating from the right or x to ∞

And if you want the central range (not a CDF): $erf(\frac{x-\mu}{\sqrt{2}\sigma})$ Integral from –x to x

Quantiles for the normal distribution

We can shift/rescale any normal to one that has μ=0 and σ=1 with an appropriate change of variables

$$t = (x - \mu)/\sigma \qquad P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \rightarrow P(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

This allows <u>quantiles</u> (one-sided from the left or right or two-sided) to be calculated in units of <u>sigma</u>

t=90%, two-sided 1.64 sigma
t=95%, two-sided 1.96 sigma

"Sigmas" as a measure of quantile

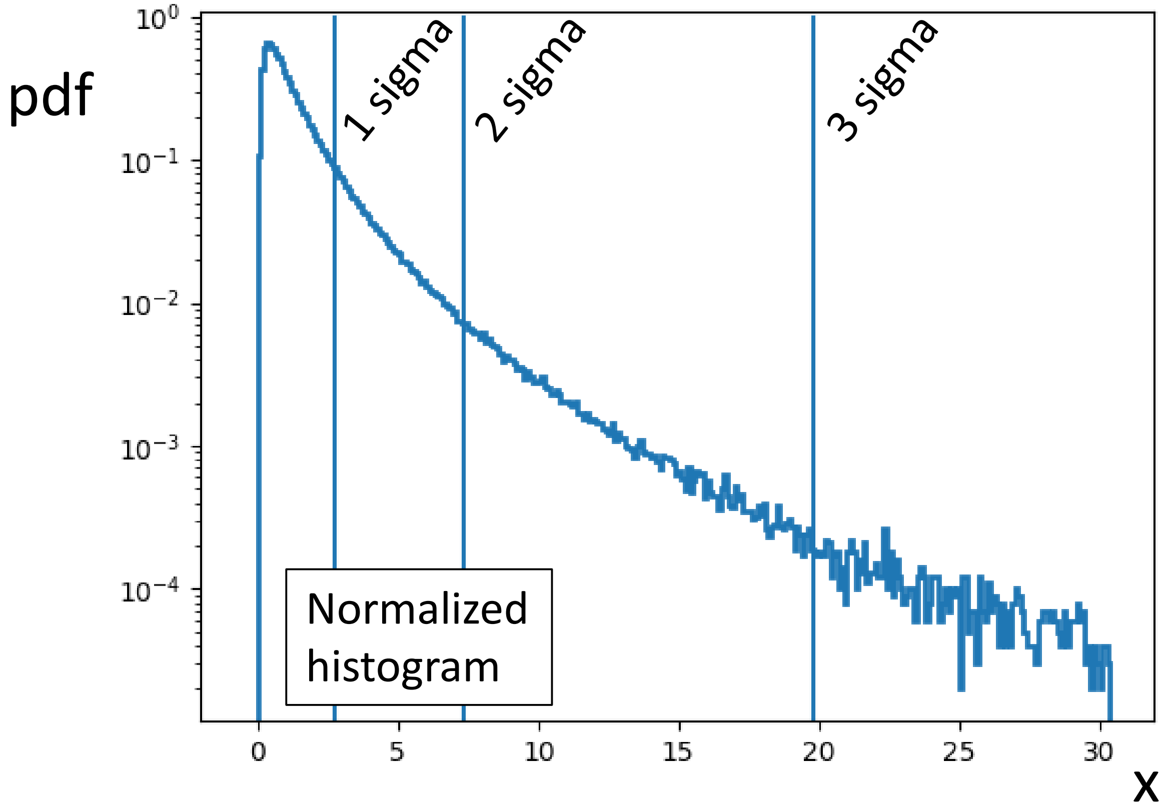
But we can flip this around and ask what is the quantile that corresponds to a given sigma.

Sigma	Quantile 2-sided	1 part in … 2-sided	Quantile 1-sided	1 part in … 1-sided
0	100%	--	50%	2
1	68.3%	3.15	84.1%	6.30
2	95.4%	22.0	97.7%	44.0
3	99.7%	370	99.9%	740
4	99.9937%	15,800	99.997%	31,600
5	99.99994%	1,740,000	99.99997%	3,490,000

68 - 95 - 99.7 rule

"Sigmas" as a measure of quantile

Sigmas are routinely used in a similar way as quantiles for distributions that are NOT normal. Here's an example with one-sided sigmas.



N.B. I've used 1,000,000 random samples from a log-normal dist.

Uniform distribution

Note that x is real
The uniform distribution is continuous:
Probability density function

Equally probable random variables between in [a,b]

Average: ½ (a + b)

Variance: $(b-a)^2 / 12$