Distributions

Monday, 14 February 2022 11:53

Gaussian distribution:

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

Mean: μ

standard deviation : σ $\rightarrow$ variance : σ^2

Cumulative density function:

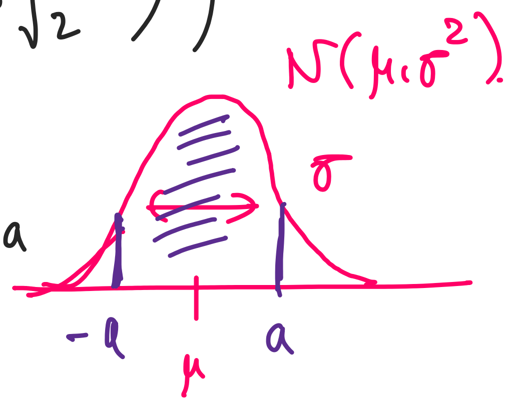
$$P(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(\frac{-(x'-\mu)^2}{2\sigma^2}\right) dx' \leftarrow$$

error function:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-t^2)\,dt.$$

$$P(x|\mu,\sigma) = \frac{1}{2}\left(1 \pm \text{erf}\left(\frac{|x-\mu|}{\sigma\sqrt{2}}\right)\right)$$

$$P(b|\mu,\sigma) - P(a|\mu,\sigma)$$



Special Case: $b = -a = \mu \pm Ma$

("$\pm Ma$" ranges around μ)

$$\text{erf}(M/\sqrt{2})$$

For $M=1 \Rightarrow 0.682 \rightarrow 1\sigma$

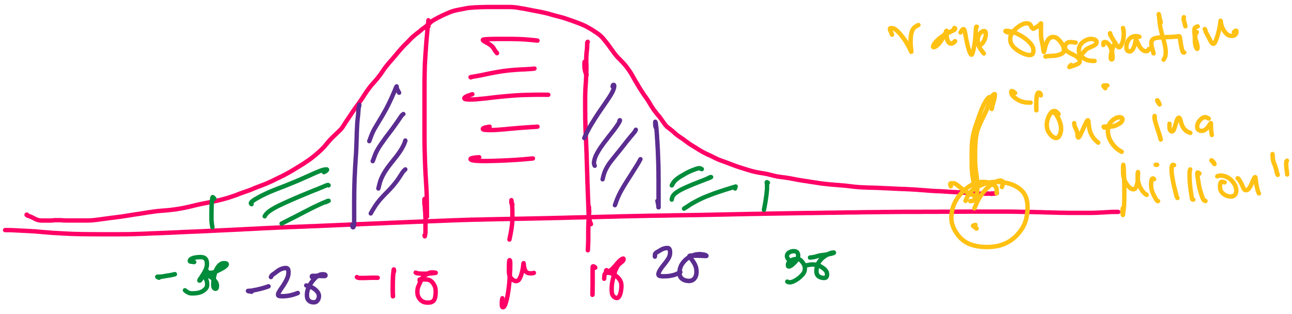
$M=2 \Rightarrow 0.954 \rightarrow 2\sigma$

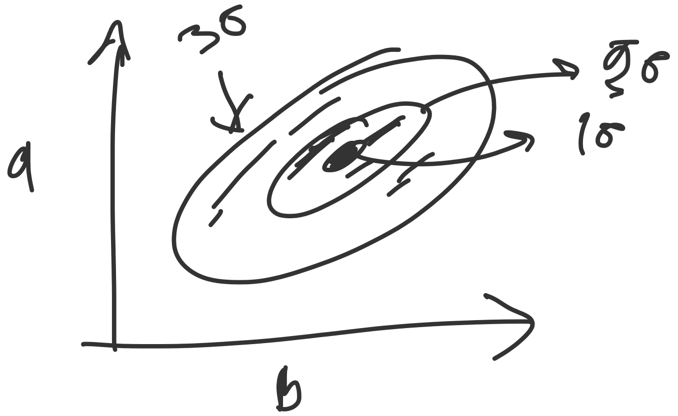
$M=3 \Rightarrow 0.997 \rightarrow 3\sigma$

"One in a Million" What is M? $\Rightarrow M = 4.9.$

OR

"One in a billion" $\Rightarrow$ M = 6.1





How about if we have outliers in the data?

- Quantiles.

interquantile range: $q_{75} - q_{25} = \sigma 2\sqrt{2}\,\mathrm{erf}^{-1}(0.5)$

$$\boxed{q_{75} - q_{25} \simeq 1.349\,\sigma.}$$

The Binomial Distribution

distribution of a variable that can take only two discrete values. (Say, (0, 1) or (success, failure) ...)

if the probability of success is 'b'.

k = how many times success occurred in N trials.

$$p(k \mid b, N) = \frac{N!}{k!\,(N-k)!}\, b^k (1-b)^{N-k}$$

Mean : $\bar{k} = bN$

Standard deviation : $\sigma_k = [N(1-b)\cdot b]^{1/2}.$

Common example of a Process following the binomial distribution function is flipping a coin.

then $b = 0.5$

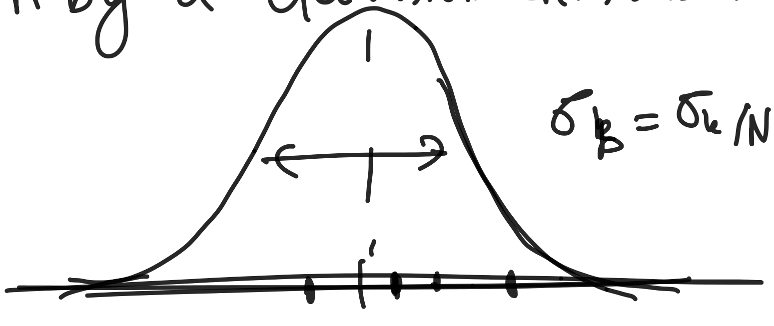
For a real coin tossed N times with k success.
What is our best estimate, $\hat{b}$, and its uncertainty given these data?

$\hat{b} = k/N \rightarrow$ uncertainty $\sigma_b \rightarrow$ data.

$(\sigma_b = \sigma_k / N)$

assume that the probability distribution for the true value of "b" is given by a Gaussian distribution

$N(\hat{b}, \sigma_b)$



$\sigma_b = \sigma_k / N$

$\mu_b = 0.5$

Check:
if $N = 10^4$
$(\hat{b} = 0.5 \pm 0.005)$ – 1% accuracy.

"How to solve this problem in a general case, without having to assume a Gaussian error distribution?"

The Poisson Distribution

– Special case of the binomial distribution.

If the number of trials N, for a binomial distribution, goes to ∞ such that the probability of success $p = K/N$, stays fixed, then the dist of the # of success is controlled by $\mu = pN$

$$P(K|\mu) = \frac{\mu^k \cdot \exp(-\mu)}{K!}$$

Mean: μ

Mode: $\mu - 1$

Standard deviation: $\sqrt{\mu}$

A μ increases

Skewness: $1/\sqrt{\mu}$

Kurtosis: $1/\mu$

both decreases as μ increases

higher order moments → Gaussian Case they vanish.

Special Case: Poisson distribution ⟶ Gaussian distributio

diff b/w the mean & the median is Not zero.

↓ q_{50} → 50^{th} quantile.

$(\text{Mean} - \text{Median}) \neq 0$ but $= 1/6$.

Sometimes the Poisson dist is also called "law of small number" OR "law of rare events".

"rare" means that only a small fraction of a large

number of trials N results in success.
("p" is small, not "μ")

Important in Astronomy:
It describes the distribution of the number of photons counted in a given interval.

Even if it is replaced by a Gaussian dist. for large μ its Poissonian origin can be recognized by $\sigma^2 = \mu$. → std. dev. $= \sqrt{\mu} = \sigma$

The Exponential (Laplacian) dist.

$$p(x|\mu,\Delta) = \frac{1}{2\Delta} \exp\left(\frac{-|x-\mu|}{\Delta}\right) \leftarrow \text{Generic}$$

this distribution is defined only for $x>0$
In this case: one-sided exponential dist.

For both $x>0$ & $x<0$, double-exponential or simply Laplacian distribution.

One-side exp. dist: $p(x|\tau) = \tau^{-1}\exp(-x/\tau)$

This distribution describes the time between two successive events which occur continuously and independently at a constant rate.

(ex. photon's arriving at a detector)

Number of such events during fixed time interval T is given by Poisson dist. with $\mu = T/\tau$.

Laplacian dist. is symmetric around μ.

Mean $= \mu$, median $= \mu$, mode $= \mu$

skewness $= 0$ (b/c its symmetric)

The standard deviation:

$$\sigma = \sqrt{2}\,\Delta \simeq 1.414\,\Delta.$$

Equivalent Gaussian width estimator from the interquartile range

$$q_{75} - q_{25} = 2\ln(2)\,\Delta$$

$$\boxed{\sigma_G = 1.028\,\Delta}$$

$\sigma_G > \sigma$ therefore this can be used to detect deviations from a Gaussian distribution towards an exponential dist

For the Gaussian dist. $|x_i - \mu| > 5\sigma$ happens in fewer than one per million. Cases.
for the exponential dist. it happen about in a thousand cases.