Lecture 3

- Descriptive statistics

↳ trying to describe the shape of arbitrary distribution $h(x)$

$\Rightarrow$ based on moments $\quad E(x^n) = \int_{-\infty}^{\infty} x^n h(x)\,dx$

e.g. $\quad$ mean $= \langle x \rangle = \int_{-\infty}^{\infty} x\,h(x)\,dx$

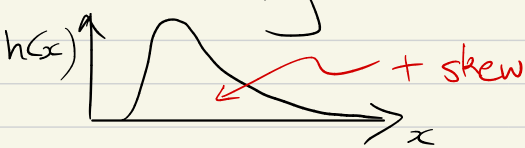
variance $= \langle (x-\mu)^2 \rangle = \int_{-\infty}^{\infty} (x-\mu)^2 h(x)\,dx$

* NOTE : moments can be biased by outliers

↳ median and σ_G better judges of location and scale of $h(x)$.

SKEWNESS = measures long tail to $+x$

3rd moment



KURTOSIS = measures "peaky-ness"

4th moment

- Sample vs. population statistics

Statistics derived from sampled data are called "sample statistics"

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

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

avoids bias, since $\bar{x}$ comes from data also.

$\boxed{\text{NOTE}}$ "s" is not the uncertainty in $\bar{x}$.
↳ it is an estimator of the scale of $h(x)$

UNCERTAINTIES : $\sigma_{\bar{x}} = \dfrac{s}{\sqrt{N}}$

$$\sigma_s = \frac{s}{\sqrt{2(N-1)}} = \frac{1}{\sqrt{2}}\sqrt{\frac{N}{N-1}}\cdot\sigma_{\bar{x}}$$

STANDARD ERROR OF QUANTILE : $\sigma_{q_p} = \dfrac{1}{h_p}\sqrt{\dfrac{p(1-p)}{N}}$

- Univariate Distribution ⇒ 1-D.

⇒ whirlwind tour through some important 1D probability distributions.

* Uniform
* Gaussian (Normal)
  ↳ confidence levels : $\pm 1\sigma = 68.3\%$
  $\pm 2\sigma = 95.4\%$
  $\pm 3\sigma = 99.7\%$

* Log normal
* χ^2 distribution
* χ^2 per dof
* Poisson distribution.
* Student's t distribution

⇒ The point is that we need to understand how SIGNIFICANT our measured statistics are, and how likely it is that they could have been made by chance noise fluctuations