The concept of Dispersion
The concept

- Law of large numbers tells us that the average result will $\rightarrow E(X)$
- What can dispersion can we expect arround $E(X)$
- Does it have a sense to use the concept of dispersion?
The concept

• Given a set of data or a functional distribution we would like to have a number for comparing dispersions.
The concept

- Given a set of data or a functional distribution we would like to have a number for comparing dispersions.

- Mean distance to the center of mass of the distribution
- Measure of curvature: Second degree polynomial
The measures

- **Standard deviation**

\[ \mu(X) = E(X) = \sum_{x \in X} xP(X = x) \]

\[ \sigma^2(X) = E(\left((X - \mu)^2\right)) = \sum_{x \in X} (x - \mu)^2 P(X = x) \]

- **Mean absolute distance**

\[ \mu(X) = E(X) = \sum_{x \in X} xP(X = x) \]

\[ MAD(X) = E(\left|X - \mu\right|) = \sum_{x \in X} |X - \mu| P(X = x) \]
The measures

- Standard deviation
- Justification:
  - Samples far from the mean increase quadratically the value of the standard deviation.
  
\[ \sigma^2(X) = E\left((X - \mu)^2\right) = \sum_{x \in X} (x - \mu)^2 P(X = x) \]

- Is the value of the curvature of the log gaussian.
The measures

• Standard deviation

• Meaning of the curvature of a plane curve:
  – Radius of a circle that approximates locally the curve.

\[ \kappa = \frac{\frac{d^2y}{dx^2}}{(1 + \left(\frac{dy}{dx}\right)^2)^{3/2}} \approx \frac{d^2y}{dx^2} \]

\[ p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

\[ \log(p(x)) = -\frac{(x-\mu)^2}{\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) \]

Radius \[= \frac{1}{d\log(p(x))} = \sigma^2 \]
The measures

- Standard deviation
- Meaning of the curvature of a plane curve:

\[ p(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} \]

\[ \log(p(x)) = -\frac{(x - \mu)^2}{\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi} \sigma}\right) \]

- Certainty=low dispersion->Small radius->small variance
- Uncertainty=high dispersion->big radius->big variance

\[ \text{Radius} = \frac{1}{\frac{d \log(p(x))}{dx}} = \sigma^2 \]
The measures

• Standard deviation
  – Certainty=low dispersion->Small radius->small variance
  – Uncertainty=high dispersion->big radius->big variance

\[ \text{Radius} = \frac{1}{d \log(p(x))} = \sigma^2 \]

\( \sigma = 0.5 \)
\( \sigma = 6 \)
Cramer-Rao and statistics
Information about Gaussian distributions.

• % of the cases for a given dispersion

\[ \mu \pm \sigma \rightarrow 68\% \]

\[ \mu \pm 2\sigma \rightarrow 95\% \]

\[ \mu \pm 3\sigma \rightarrow 99\% \]
The measures

• Mean absolute distance

\[ \mu(X) = E(X) = \sum_{x \in X} x P(X = x) \]

\[ MAD(X) = E(|X - \mu|) = \sum_{x \in X} |X - \mu| P(X = x) \]

Samples far from the mean increase linearly the value of the standard deviation. Note: outliers do not count as much as in the standard deviation.
The measures

• Mean absolute distance
  – Uses: measure of dispersion for distributions with longer tail than gaussians.

\[
MAD(X) = E(|X - \mu|) = \sum |X - \mu|P(X = x)
\]

Laplace vs. Gaussian
Note: That high values are much more probable in the case of an exponential random variable
The measures

• Mean absolute distance
  – Outliers do not change so much the value of the estimation.

\[
MAD(X) = E\left( |X - \mu| \right) = \sum_{x \in X} |X - \mu| P(X = x)
\]
The concept

• Given a set of data or a functional distribution we would like to have a number for comparing dispersions.
The Box Plot

• Gives information about the dispersion summarizing the information about:
  – The interquantile size. $x_{25} - x_{75}$
  – Median
  – Sample nearest to the 1.5 times the interquantile margin
  – The outliers (points $>1.5$ IQR)

• Examples:
  100 points
  Gaussian vs. geometric
Description of a random variable

• Gaussian

\[ p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} \]
Description of a random variable

- Geometric: 
  \[ P(X = i) = p^{i-1}(1 - p) \quad \text{for } i = 1, 2, 3, \ldots \]
Description of a random variable

• Negative binomial
Description of a random variable

- Poisson
Some intuitions

• The **flaw** of averages

**Markowitz's Idea:**
• Introduce variability when assessing the value of an asset.
• Maximize mean, minimizing variance.

What is better?:
A-Mean benefit of 500 plus minus 400
B-Mean benefit of 200 plus minus 50

http://www.stanford.edu/~esavage/flaw/Article.htm
The flaw of averages

• An investment problem
  – Suppose you want your $200,000 retirement fund invested in the Standard & Poor's 500 index to last 20 years. How much can you withdraw per year?
  – The return of the S&P has varied over the years but has averaged about 14 percent per year since 1952.
  – If you do this you will be pleased to find that you can withdraw $32,000 per year.

\[
A = 200.000
\]
\[
r = 14\%
\]
\[
x = \frac{rA(1+r)^{20}}{(1+r)^{20} - 1}
\]

\[
A(1+r)^{20} - \sum_{k=0}^{19} x(1+r)^k = 0
\]
The flaw of averages

- Model of the return:
  - \( r \% \) with probability \( p \)
  - Histogram of the return
  - Average value \( r \), but can fluctuate.
    - Sometimes gives benefits
    - Sometimes losses
  - Note that each month a fixed quantity is subtracted independently of \( r \)
The flaw of averages

- Simulations on real data:

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>14% Tanks in 8 yrs.</td>
<td>15.4% Goes the distance.</td>
</tr>
<tr>
<td>15.4% Tanks in 13 yrs.</td>
<td>15.3% Tanks in 10 yrs</td>
</tr>
</tbody>
</table>

The Flaw of Averages
BY SAM SAVAGE
Published Sunday, October 8, 2000, in the San Jose Mercury News
The flaw of averages

• Model of the return:
  – $r\%$ with probability $p$
  – $(1+f)r\%$ with probability $(1-p)/2$
  – $(1-f)r\%$ with probability $(1-p)/2$

• Simulation.
  – 4.000.000 runs.

Taken from Tijms, Understanding probability
Tchebychev Inequality

• A bound on the upper probability.

\[ P\left(\left| X - E(X) \right| > c \right) \leq \frac{\sigma^2}{c^2} \]
Tchebychev Inequality

- Geometrical meaning of

\[ P\{X - E(X) > c\} \leq \frac{\sigma^2}{c^2} \]

\[ f(c) \propto \frac{1}{c^2} \]
Tchebychev Inequality

• Geometrical meaning of \[ P\{X - E(X) > c \} \leq \frac{\sigma^2}{c^2} \]

• What happens with?

\[
P(X = x) = \frac{1}{\pi(1+x^2)}
\]

\[
P(X = x) = \frac{1}{x^\alpha}
\]

Only valid on distributions that have finite variance!
Tchebychev Inequality

• For a given probability distribution of a random variable $X$, with finite variance we have:

$$P\{|X - E(X)| > k\sigma\} \leq \frac{1}{k^2}$$

• for any $k>0$ or equivalently

$$P\{|X - E(X)| > c\} \leq \frac{\sigma^2}{c^2}$$
Tchebychev Inequality

• Proof
  – Given an ordered set \( \{x_1, x_2, x_3, x_4, x_5, \ldots \} \)
  – We define the subset \( A = \{k \mid |x_k - E(X)| > c \} \)
  – then

\[
\sigma^2 = \sum_{k} (x_k - E(X))^2 p_k \geq \sum_{k \in A} (x_k - E(X))^2 p_k \geq c^2 \sum_{k \in A} p_k
\]

\[
\sigma^2 \geq c^2 P\{|X - E(X)| > c \}
\]
Tchebychev Inequality

• Note that the inequality can be rough, and highly inexact for high values of $c$

• Uses:
  – Information theory, bounds on probabilities and events highly unprobable.

\[ \frac{1}{c^2} 
\]
Random variables without variance

• Family known as
  – Pareto Stable or Mandelbrot Levy

• Models:
  – Internet traffic
  – Processes in unix systems
  – Speculative prices/Pluviometric Data
Speculative Prices

• Mandelbrot’s paper on long tail densities
  – An interesting result

http://classes.yale.edu/fractals/Panorama/ManuFractals/Internet/Internet4.html
Syndrom of infinite variance

• Mandelbrot’s paper on long tail densities

\[
\mu(X) = E(X) = \sum_{x \in X} xP(X = x)
\]

\[
\sigma^2(X) = E\left((X - \mu)^2\right) = \sum_{x \in X} (x - \mu)^2 P(X = x)
\]

\[
P(X = x) = \frac{1}{x^\alpha}
\]

Note that for \(\alpha<2\) the sum diverges.
\[ p(x) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]