

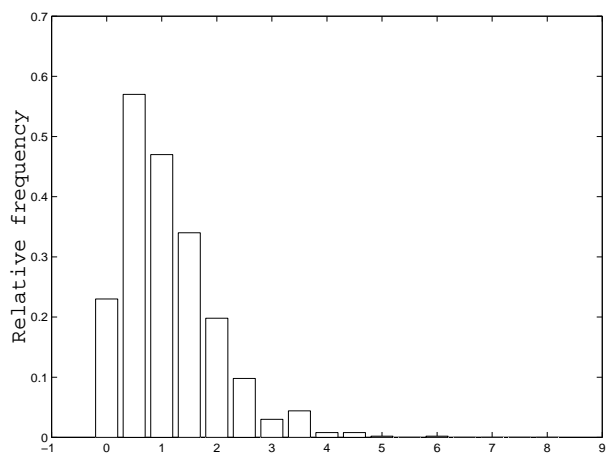
Tchebysheff's Theorem Given a number $k \geq 1$ and a population with n measurements, at least $[1 - (1/k^2)]$ of the measurements will lie within k standard deviations of their mean.

A Simplified Tchebysheff's Theorem

At least $3/4$ of th measurements lie in the interval $(\mu - 2\sigma, \mu + 2\sigma)$.

At least $8/9$ of th measurements lie in the interval $(\mu - 3\sigma, \mu + 3\sigma)$.

The Tchebysheff's Theorem applies to any shape of relative frequency histogram (r.f.h.), so it is very conservative. The following r.f.h. is not symmetric (skewed).



Example 1.4

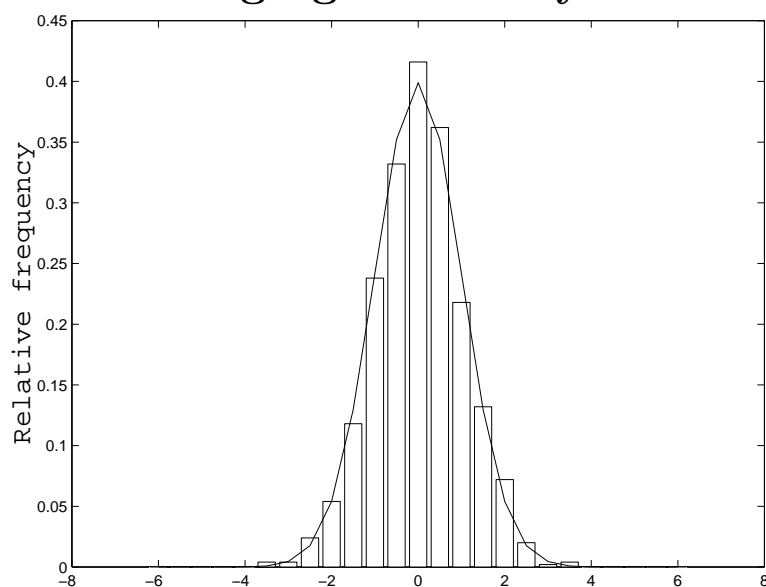
Suppose that the population histogram is symmetric about its mean and the mean and variance of the population of $n = 108$ measurements are 60 and 100, respectively. Use Tchebysheff's theorem to prove that at most 13 measurements are greater than 80.

Proof: In Tchebysheff's theorem, we take $k = 2$, then at least $3/4$ of the measurements lie within $[60 - 2 * 10, 60 + 2 * 10]$. In other words, at most $1/4$ or **27** of the measurements lie outside of $[60 - 2 * 10, 60 + 2 * 10]$. Since the sample histogram is symmetric about its mean, at most **13** of the measurements are greater than $60 + 2 * 10 = 80$.

If the shape of the r.f.h. is symmetric and bell shaped, (This is just the normal distribution), then we have a better estimation rule. Precisely, if the r.f.h broken line is very close to the curve of function:

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

then we say that the r.f.h. is symmetric and bell shaped. The following figure show you this meaning.



For a symmetric and bell shaped r.f.h., we have following better estimation rule.

Empirical Rule

Given a population of measurements that is approximately bell shaped, then we have the following estimations:

The interval $(\mu - \sigma, \mu + \sigma)$ contains approximately 68% of the measurements.

The interval $(\mu - 2\sigma, \mu + 2\sigma)$ contains approximately 95% of the measurements.

The interval $(\mu - 3\sigma, \mu + 3\sigma)$ contains all or almost all of the measurements.

Example 1.5

The number of television viewing hours per household and the prime viewing times are two factors that affect television advertising income. A random sample of 25 households in a particular viewing area produced the following estimates of viewing hours per household:

6.5	8.0	4.0	5.5	6.0
3.0	6.0	7.5	15.0	12.0
5.0	12.0	1.0	3.5	3.0
7.5	5.0	10.0	8.0	3.5
9.0	2.0	6.5	1.0	5.0

Find the percentage of the viewing hours per household that falls into the interval $(\bar{x} - 2s, \bar{x} + 2s)$. Compare with the corresponding percentage given by the Empirical Rule.

Solution Here $n = 25$, $\sum_{i=1}^{25} x_i = 155.5$, $\sum_{i=1}^{25} x_i^2 = 1260.75$.
Then

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{25} \sum_{i=1}^{25} x_i = \frac{155.5}{25} = 6.22 \\ s &= \sqrt{\frac{\sum_{i=1}^n \{x_i - \bar{x}\}^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{25} x_i^2 - \frac{(\sum_{i=1}^{25} x_i)^2}{n}}{n-1}} \\ &= \sqrt{\frac{1260.75 - \frac{(155.5)^2}{25}}{24}} = 3.497\end{aligned}$$

We can get $(\bar{x} - 2s, \bar{x} + 2s) = (6.22 - 6.994, 6.22 + 6.994) = (-0.774, 13.214)$. From the original data, 24 measurements or $(24/25)100 = 96\%$ of the measurements fall in this interval. This is close to the 95% result of Empirical rule.

Definition 4.1 If an experiment can be repeated under the same condition, its outcome cannot be predicted with certainty, and the collection of its every possible outcome can be described prior to its performance, this kind of experiment is called random experiment. All the possible outcomes of a random experiment is denoted by Ω which is called sample space. A sub-

set of Ω is called an event. An event cannot be decomposed is called a simple event. Two events are mutually exclusive if there are no intersections.

Example 4.1

Experiment: Roll a fair die on a hard, flat floor and observe the number appearing on the upper face. Is this a random experiment? If yes, what is the sample space? Define $E_i = \{\text{Observe a } i\}, i = 1, \dots, 6$, $A = \{\text{Observe an odd number}\}$, and $B = \{\text{Observe a number less than or equal to 4}\}$. Is A an event? Is A a simple event? Do A and B be mutually exclusive?

Classical Probability

Definition 4.2 Given a sample space $\Omega = \{E_1, E_2, \dots, E_n\}$ with finite simple events, if we define a function \mathbb{P} on Ω by:

- (1) The empty set is denoted by ϕ . $\mathbb{P}(\phi) = 0$.
- (2) $\mathbb{P}(E_i) = 1/n$ for each $1 \leq i \leq n$. This means that each simple event is equally likely.
- (3) For each event A , $\mathbb{P}(A)$ is equal to the sum of the probabilities of simple events contained in A .

Then, \mathbb{P} is called a classical probability on Ω .

Remark:

- (1) $0 \leq \mathbb{P}(A) \leq 1$.
- (2) $\mathbb{P}(\Omega) = 1$.

A more general definition of probability will be introduced later on.

Example 4.2 Consider the random experiment and events defined in Example 4.1, where

$A = \{\text{Observe an odd number}\}$, and

$B = \{\text{Observe a number less than or equal to 4}\}$. Find the classical probability of event A and B .

Solution:

$$\mathbb{P}(A) = \mathbb{P}(E_1) + \mathbb{P}(E_3) + \mathbb{P}(E_5) = 1/6 + 1/6 + 1/6 = 1/2$$

$$\mathbb{P}(B) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) + \mathbb{P}(E_4) = 1/6 + 1/6 + 1/6 + 1/6 = 2/3$$

Example 4.3 A jar contains four coins: a nickel, a dime, a quarter, and a half-dollar. Three coins are randomly selected from the jar.

- List all the simple events in the sample space Ω .
- This is a typical example of the classical probability. What is the probability that the selection will contain the half-dollar?
- What is the probability that the total amount drawn will equal 0.6 dollar or more?

Solution:

a. Denote:

N: nickel;

D: dime;

Q: quarter;

H: half-dollar.

and $E_1 = (NDQ), E_2 = (NDH), E_3 = (NQH), E_4 = (DQH)$.

Then, $\Omega = \{E_1, E_2, E_3, E_4\}$.

b.

$$\mathbb{P}(\text{choose a half-dollar}) = \mathbb{P}(E_2) + \mathbb{P}(E_3) + \mathbb{P}(E_4) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

c. The simple event along with their monetary values follow:

$$E_1 = NDQ = \$0.4$$

$$E_2 = NDH = \$0.65$$

$$E_3 = NQH = \$0.80$$

$$E_4 = DQH = \$0.85$$

Hence,

$$\mathbb{P}(\text{total amount is } \$0.6 \text{ or more}) = \mathbb{P}(E_2) + \mathbb{P}(E_3) + \mathbb{P}(E_4) = \frac{3}{4}.$$