Basics of probability

IC152 Feb 2021

Experiments: any process that produces an outcome

Outcomes: An observation

Random experiment: outcome is not predictable

Sample space: set of all outcomes

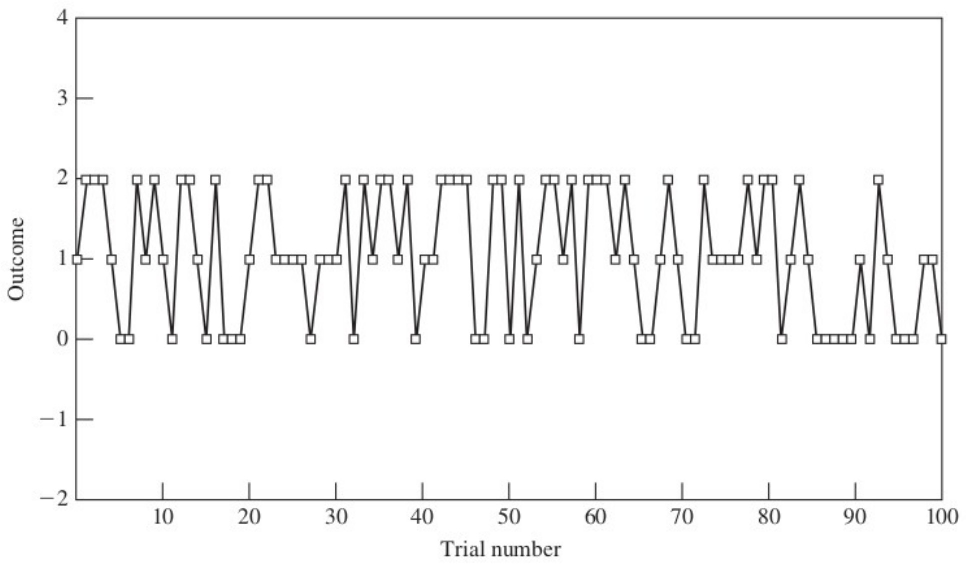
Event: A set of outcomes

Example: Tossing of two coins and noting if they are head or tail.
S = {(HH),(HT),(TH),(TT)}: Sample space
A = {(HH),(HT)}: A is the event that the first coin lands on heads
B = {(HT),(TT)}: B is the event that the second coin lands on tails
$A \cap B$ = {(HT)}: event that the first coin landed heads and second landed tails

Another example:
Urn and three balls marked 0,1,2.
Urn shaken, ball picked, number noted.

What is the outcome of this experiment?

Number from the set $S = \{0, 1, 2\}$.



Probability models depend on
statistical regularity.

Averages obtained in long sequences of
repetitions yield the same value.

From "Probability, Statistics and Random Processes for Electrical Engineering", by Leon Garcia, 3rd Ed.

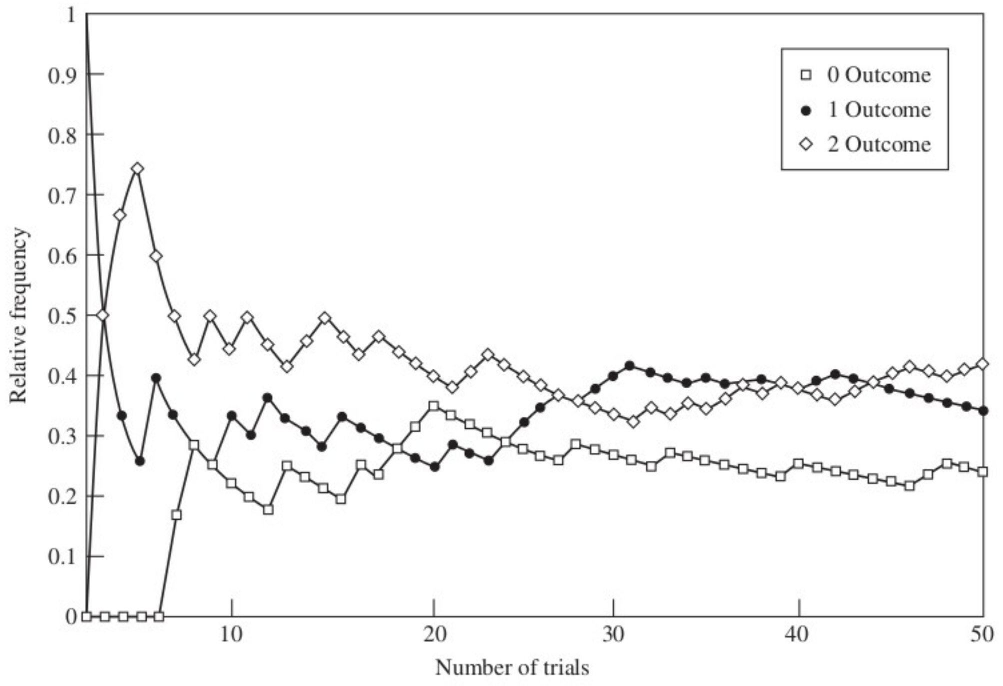
- Repeat the urn experiment n times under identical conditions.
- Number of times ball k appears: $N_k(n)$, $k = \{0, 1, 2\}$
- **Relative frequency** of outcome k:

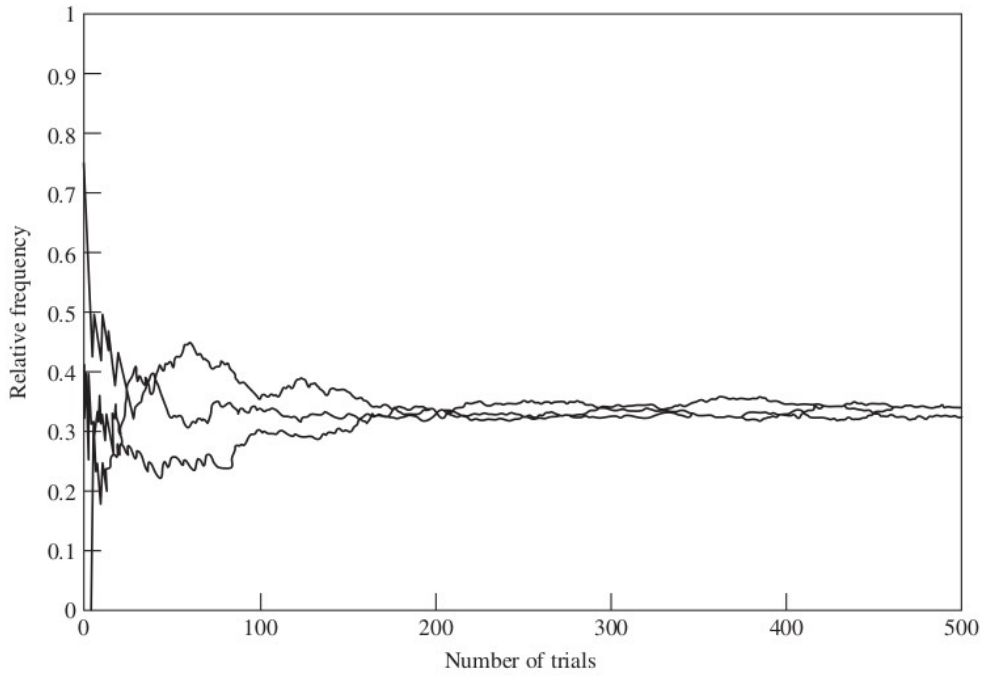
$$f_k(n) = \frac{N_k(n)}{n}$$

- We have

$$\lim_{n \to \infty} f_k(n) = p_k$$

where p_k is the probability of outcome k.





For more number of trials

- Consider n trials of a random experiment with K possible outcomes.
- $S = \{1, 2, \ldots K\}$
- We have

$$0 \leq N_k(n) \leq n \quad \text{for } k = 1, 2 \ldots K$$

- This means relative frequencies are between zero and one:

$$0 \leq f_k(n) \leq 1 \quad \text{for } k = 1, 2 \ldots K$$

- Also,

$$\sum_{k=1}^{K} N_k(n) = n$$

and

$$\sum_{k=1}^{K} f_k(n) = 1$$

- Let C be the event "A or B occurs," where A and B are events that cannot occur simultaneously. The number of times C occurs is $N_C(n) = N_A(n) + N_B(n)$. Also,

$$f_C(n) = f_A(n) + f_B(n)$$

Axiomatic approach to probability

- Probability was defined by its long-term relative frequency
- Several problems with this definition:
  - Limit may not be defined
  - We cannot perform an experiment infinite times, therefor p_k is an approximation
  - What about experiments that cant be repeated?
- Thus, we need another definition, independent of the application
- Also, we must be able to interpret probability intuitively as relative frequency

Let us assume:

- A random experiment has been defined
- The set of all possible outcomes S has been defined
- Subsets of S called events have been defined
- Each event has been assigned a number P[A], such that:

- $0 \leq P[A] \leq 1$
- $P[S] = 1$
- If A and B cannot occur simultaneously, then $P[A \vee B] = P[A] + P[B]$.

  - We are not worried about what the probabilities mean, or how they are obtained.
  - This depends on the user or application

- Telephone conversation: is the speaker talking or is silent?
- Typical speaker only active 1/3rd of the time (someone has to give this to you; eg. a domain expert.)

- Random experiment:
  - Urn with 2 white balls (silence) and 1 black ball (speech)

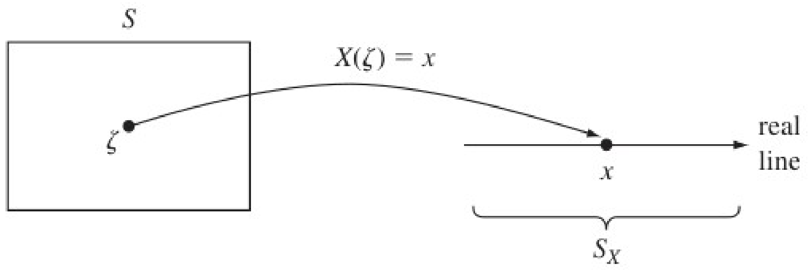
- *Question:* what is the probability that more than 24 out of 48 independent speakers are active at the same time?
- *Model:* what is the probability that 24 black balls are selected in 48 independent repetitions of the urn experiment?

Random experiments and event classes

- Sample space S: all possible outcomes of the random experiment
- Define the event class $\mathcal{F}$ of all events of interest
- Events in $\mathcal{F}$ are assigned probabilities
- $\mathcal{F}$ is closed under complements, countable unions and intersections
- When S is finite or countable, $\mathcal{F}$ is the powerset of S.
- When S is uncountable (eg. $\mathbb{R}$), F cannot be the set of all subsets of S (it will not satisfy all axioms of probability)
- We can model events of practical interest by countable unions and intersections of intervals of $\mathbb{R}$.

Random variables

- A random variable is a function that assigns a real number to each outcome in the sample space.



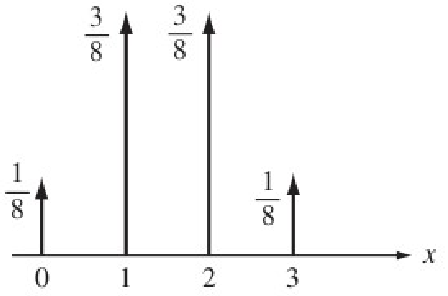
S is the domain
S_X is the range

Discrete random variables

- For a discrete random variable, the range S_X is a countable set.
- The probability mass function (PMF) is:

$$P_X(x) = P[X = x] = P[\{\zeta : X(\zeta) = x\}] \ \text{for } x \in \mathbb{R}$$

- If $S_X = \{x_1, x_2 \ldots, x_n\}$, then let $p_i = P[X = x_i] = P(x_i)$



What random variable does this PMF represent?

Number of heads in three independent tosses of a coin

Number of sweets eaten in a pack of three sweets

Expected value

- The expected value of X is

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

- The variance is

$$\sigma^2 = E[(x-\mu)^2] = \sum_{x \in S_X} (x-\mu)^2 P(x)$$

Contrast this with the sample mean and sample variance

Pairs of discrete random variables

- Two random variables X and Y taking values from $S_X = \{v_1, v_2, \ldots, v_m\}$ and $S_Y = \{w_1, w_2 \ldots, w_n\}$
- (x, y) is a point in $S_X \times S_Y$
- Joint probability $p_{i,j} = P[x = v_i, y = w_j]$
- Marginal distributions for X and Y

$$P_X(x) = \sum_{y \in S_Y} P(x, y)$$

$$P_Y(y) = \sum_{x \in S_X} P(x, y)$$

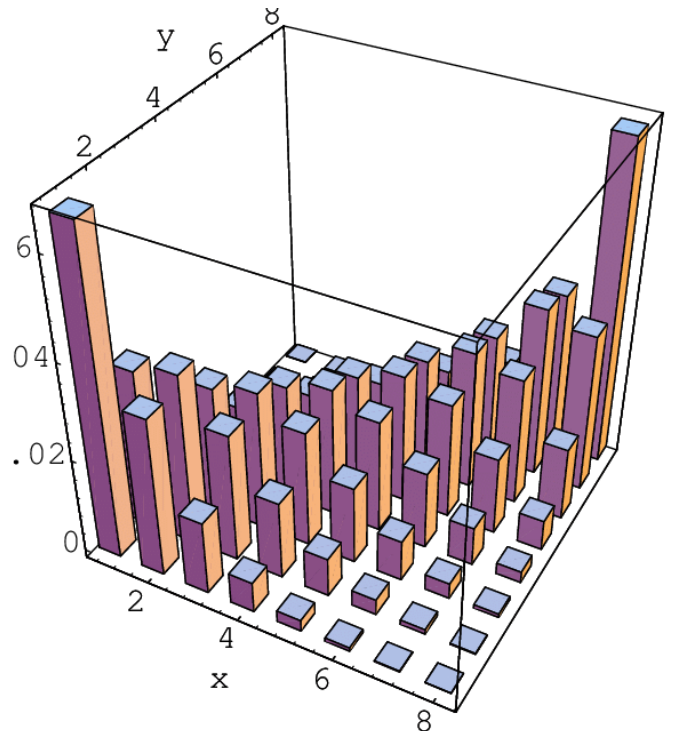


Figure from researchgate.net

Joint distributions and conditional distributions

- $N = 60$ samples drawn from joint distribution (top left figure.) X takes 9 possible values and Y takes 2 values.

- Top right: Histogram estimate of marginal distribution of X. Bottom left: Histogram estimate of marginal distribution of Y.

- Bottom right: The conditional distribution $p(X|Y = 1)$

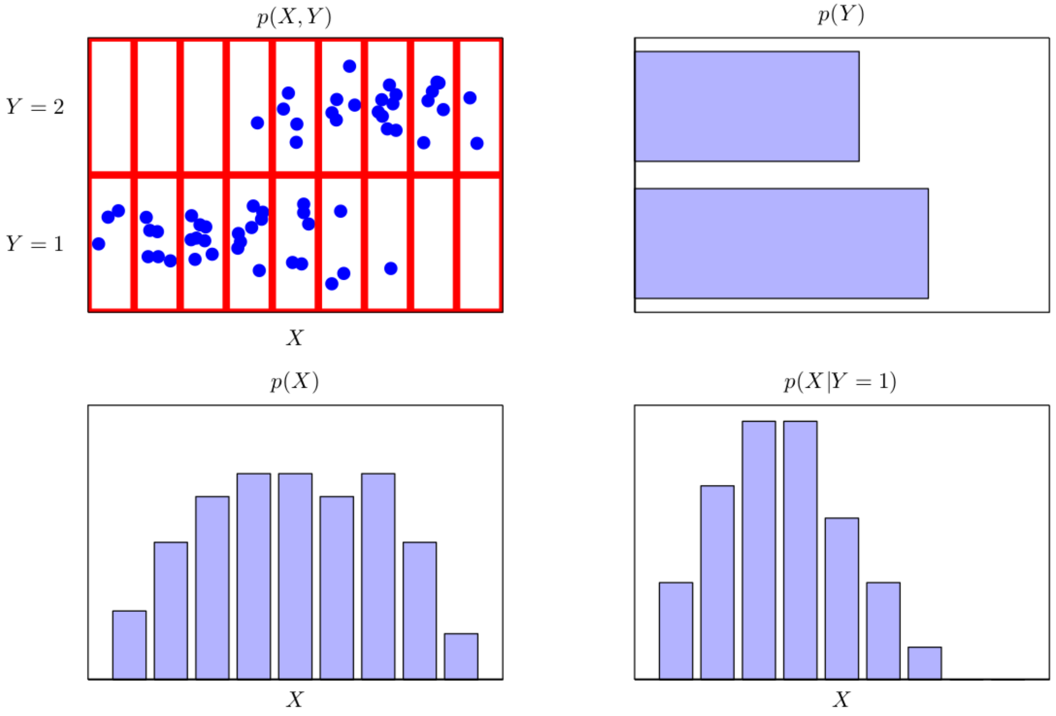


Figure from Pattern Recognition and Machine Learning, C.M. Bishop

Independence

Two random variables X and Y are independent iff

$$P(x, y) = P_X(x)P_Y(y)$$

- Let $p_i = P(X = v_i)$ and let $q_j = P(Y = w_j)$. p_i is the fraction of time that $X = v_i$, and q_j is the fraction of time that $Y = w_j$.
- Consider the situation when $x = v_i$.
- If it is true that the fraction of situations in which $y = w_j$ is still q_j, then knowing the value of X did not give us any additional knowledge about Y. Then X and Y are independent.

Example of independent events.

E1: Mark Zukerberg drives at more than 60 mph on his office commute.

E2: I have more than three rotis for dinner.

Example of events that need not be independent.

E1: You go to a shopping mall with a multiplex.

E2: You eat popcorn.

Two rules of probability

Sum rule

$$P(x) = \sum_{Y} P(x, y)$$

Product rule

$$P(x, y) = P(y|x)P(x)$$

Bayes rule

Using the fact that $P(x, y) = P(y, x)$, and the sum and product rules, we get **Baye's rule**:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

HW: Using the sum rule and product rule, derive Baye's rule.

Binomial random variable

$$S_X = \{0, 1, 2, \ldots n\}$$

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \ldots n$$

$$E[X] = np \quad \mathrm{var}[X] = np(1-p)$$

X is the number of successes in n iid trials, with the probability of success in each trial being p.

Poisson random variable

$$S_X = 0, 1, 2, \ldots$$

$$p_k = \frac{\alpha^k}{k!} e^{-\alpha} \quad k = 0, 1, \ldots \;\; \text{and} \;\; \alpha > 0$$

$$E[X] = \alpha \quad \text{var}[X] = \alpha$$

X is the number of events that occur in one time unit when time between events is exponentially distributed with mean $1/\alpha$.

Uniform random variable

$$S_X = 0, 1, 2, \ldots, L$$

$$p_k = \frac{1}{L} \quad k = 0, 1, \ldots, L$$

Demo code for sampling data from various distributions and plotting their histograms

```
from numpy.random import default_rng
import numpy as np
import matplotlib.pyplot as plt

rng = default_rng()

##################################
# binomial distribution
##################################
n, p = 20, 0.75 # number of trials, probability of each trial
# result of flipping a coin 20 times, tested 1000 times.
s = rng.binomial(n, p, 1000)

# plot a (not so good) histogram
#nh,bh,ph = plt.hist(s,bins=20)

# plot a better histogram with custom bins and xticks
myBins = np.arange(0,21,1)
nh,bh,ph = plt.hist(s,bins=myBins)
plt.xticks(myBins)


##################################
# poisson distribution
##################################
#s = rng.poisson(25, 10000)
#myBins = np.arange(0,50,1)
#nh,bh,ph = plt.hist(s,bins=myBins)
#plt.xticks(myBins)


##################################
# uniform distribution
##################################
# s = rng.integers(51, size=10000)
# myBins = np.arange(-5,55,1)
# nh,bh,ph = plt.hist(s,bins=myBins)
# plt.xticks(myBins)
```