확률과 통계

Class 2

Chapter 3
Discrete Random Variables and Their Distributions

3.1 Distribution of a random variable

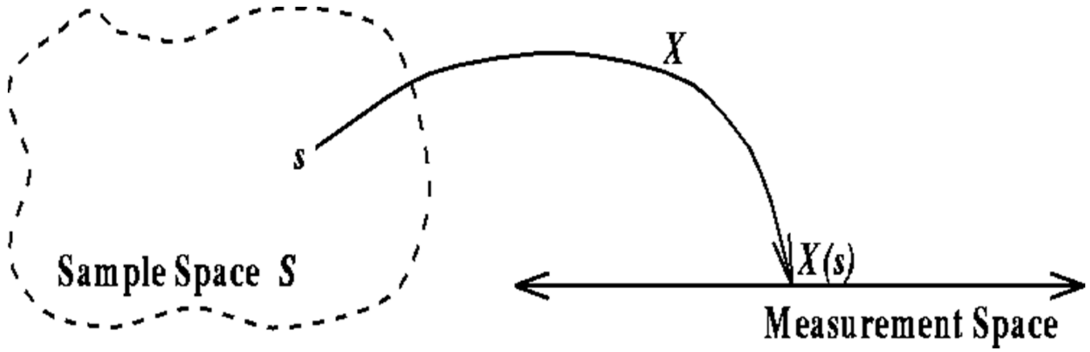
3.1.1 Main concepts

DEFINITION 3.1

A **random variable** is a function of an outcome,

$$X = f(\omega).$$

In other words, it is a quantity that depends on chance.



Example 3.1

- Consider an experiment of tossing 3 fair coins and counting the number of heads.
- Certainly, the same model suits the number of girls in a family with 3 children,
  the number of 1's in a random binary code consisting of 3 characters, etc.
- Let X be the number of heads (girls, 1's).
- Prior to an experiment, its value is not known.
- All we can say is that X has to be an integer between 0 and 3.

- Since assuming each value is an event, we can compute probabilities,

$$\boldsymbol{P}\{X=0\} = \boldsymbol{P}\{\text{three tails}\} = \boldsymbol{P}\{TTT\} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$
$$\boldsymbol{P}\{X=1\} = \boldsymbol{P}\{HTT\} + \boldsymbol{P}\{THT\} + \boldsymbol{P}\{TTH\} = \frac{3}{8}$$
$$\boldsymbol{P}\{X=2\} = \boldsymbol{P}\{HHT\} + \boldsymbol{P}\{HTH\} + \boldsymbol{P}\{THH\} = \frac{3}{8}$$
$$\boldsymbol{P}\{X=3\} = \boldsymbol{P}\{HHH\} = \frac{1}{8}$$

- Summarizing,

x	$\boldsymbol{P}\{X=x\}$
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

DEFINITION 3.2

Collection of all the probabilities related to X is the **distribution** of X. The function

$$P(x) = \boldsymbol{P}\{X = x\}$$

is the **probability mass function**, or **pmf**. The **cumulative distribution function**, or **cdf** is defined as

$$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} \boldsymbol{P}(y). \tag{3.1}$$

The set of possible values of X is called the **support** of the distribution F.

$$\sum_x P(x) = \sum_x \boldsymbol{P}\{X = x\} = 1. \qquad \lim_{x \downarrow -\infty} F(x) = 0 \text{ and } \lim_{x \uparrow +\infty} F(x) = 1.$$

x	$\boldsymbol{P}\{X = x\}$
0	$1/8$
1	$3/8$
2	$3/8$
3	$1/8$
Total	1

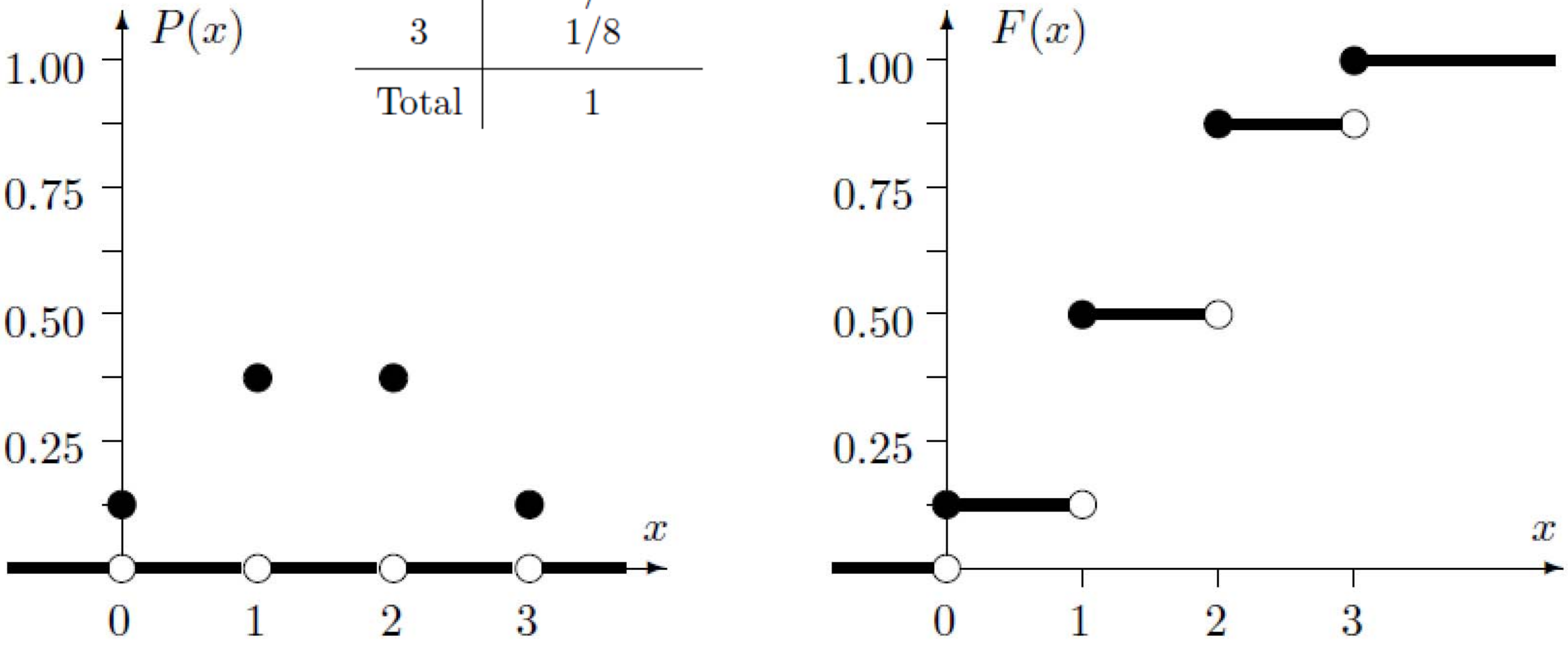
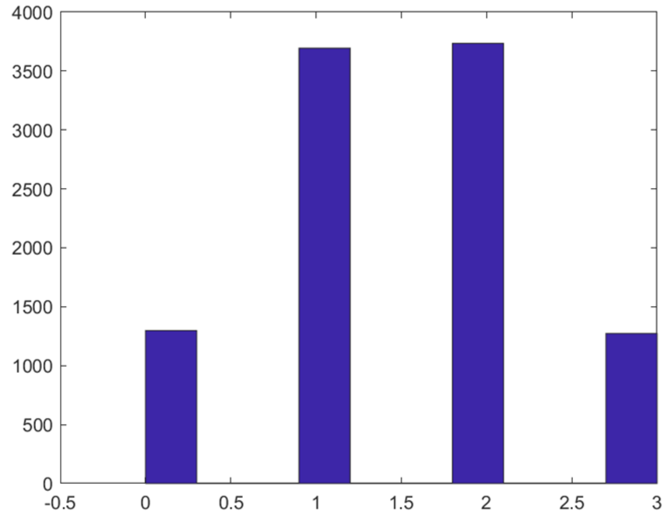


FIGURE 3.1: *The probability mass function $P(x)$ and the cumulative distribution function $F(x)$ for Example 3.1. White circles denote excluded points.*

Matlab Demo

- The following MATLAB code simulates 3 coin tosses and computes a value of X 10,000 times.
  - `N = 10000; % Number of simulations`
  - `U = rand(3,N); % a 3-by-N matrix of random numbers from [0,1]`
  - `Y = (U < 0.5); % Y=1 (heads) if U < 0.5, otherwise Y=0 (tails)`
  - `X = sum(Y); % Sums across columns. X = number of heads`
  - `hist(X); % Histogram of X`



Example 3.3 (Errors in independent modules)

- A program consists of two modules.
- The number of errors X_1 in the first module has the pmf $P_1(x)$, and the number of errors X_2 in the second module has the pmf $P_2(x)$, independently of X_1, where

x	$P_1(x)$	$P_2(x)$
0	0.5	0.7
1	0.3	0.2
2	0.1	0.1
3	0.1	0

- Find the pmf and cdf of $Y = X_1 + X_2$, the total number of errors.
- Solution.
  - We break the problem into steps.
  - First, determine all possible values of Y, then compute the probability of each value.
  - Clearly, the number of errors Y is an integer that can be as low as $0 + 0 = 0$ and as high as $3 + 2 = 5$.
  - Since $P_2(3) = 0$, the second module has at most 2 errors.

- Next,

$$P_Y(0) = P\{Y = 0\} = \boldsymbol{P}\{X_1 = X_2 = 0\} = P_1(0)P_2(0) = (0.5)(0.7) = 0.35$$
$$P_Y(1) = P\{Y = 1\} = P_1(0)P_2(1) + P_1(1)P_2(0) = (0.5)(0.2) + (0.3)(0.7) = 0.31$$
$$P_Y(2) = P\{Y = 2\} = P_1(0)P_2(2) + P_1(1)P_2(1) + P_1(2)P_2(0) = (0.5)(0.1) + (0.3)(0.2) + (0.1)(0.7) = 0.18$$
$$P_Y(3) = P\{Y = 3\} = P_1(1)P_2(2) + P_1(2)P_2(1) + P_1(3)P_2(0) = (0.3)(0.1) + (0.1)(0.2) + (0.1)(0.7) = 0.12$$
$$P_Y(4) = P\{Y = 4\} = P_1(2)P_2(2) + P_1(3)P_2(1) = (0.1)(0.1) + (0.1)(0.2) = 0.03$$
$$P_Y(5) = P\{Y = 5\} = P_1(3)P_2(2) = (0.1)(0.1) = 0.01$$

- Now check:

$$\sum_{y=0}^{5} P_Y(\mathrm{y}) = 0.35 + 0.31 + 0.18 + 0.12 + 0.03 + 0.01 = 1$$

thus we probably counted all the possibilities and did not miss any.

- The cumulative function can be computed as

$$F_Y(0) = P_Y(0) = 0.35$$
$$F_Y(1) = F_Y(0) + P_Y(1) = 0.35 + 0.31 = 0.66$$
$$F_Y(2) = F_Y(1) + P_Y(2) = 0.66 + 0.18 = 0.84$$
$$F_Y(3) = F_Y(2) + P_Y(3) = 0.84 + 0.12 = 0.96$$
$$F_Y(4) = F_Y(3) + P_Y(4) = 0.96 + 0.03 = 0.99$$
$$F_Y(5) = F_Y(4) + P_Y(5) = 0.99 + 0.01 = 1.00$$

- Between the values of Y , $F(x)$ is constant.

3.1.2 Types of random variables

- So far, we are dealing with **discrete random variables**.
- These are variables whose range is finite or countable.
- In particular, it means that their values can be listed, or arranged in a sequence.
- Examples include the number of jobs submitted to a printer,
  the number of errors,
  the number of error-free modules,
  the number of failed components, and so on.
- Discrete variables don't have to be integers.
- For example, the proportion of defective components in a lot of 100 can be 0, 1/100, 2/100, ..., 99/100, or 1.
- This variable assumes 101 different values,
  so it is discrete, although not an integer.
- On the contrary, **continuous random variables** assume a whole interval of values.
- This could be a bounded interval (a, b),
  or an unbounded interval such as $(a,+\infty)$, $(-\infty, b)$, or $(-\infty,+\infty)$.
- Sometimes, it may be a union of several such intervals.
- Intervals are uncountable, therefore,
  all values of a random variable cannot be listed in this case.
- Examples of continuous variables include various times (software installation time, code execution time, connection time, waiting time, lifetime),
  also physical variables like weight, height, voltage, temperature, distance, the number of miles per gallon, etc.

Example 3.4

- For comparison, observe that a long jump is formally a continuous random variable because an athlete can jump any distance within some range.
- Results of a high jump, however, are discrete because the bar can only be placed on a finite number of heights.

Example 3.5

- A job is sent to a printer.
- Let X be the waiting time before the job starts printing.
- With some probability, this job appears first in line and starts printing immediately, $X = 0$.
- It is also possible that the job is first in line but it takes 20 seconds for the printer to warm up, in which case $X = 20$.
- So far, the variable has a discrete behavior with a positive pmf $P(x)$ at $x = 0$ and $x = 20$.
- However, if there are other jobs in a queue, then X depends on the time it takes to print them, which is a continuous random variable.
- Using a popular jargon, besides "point masses" at $x = 0$ and $x = 20$, the variable is continuous, taking values in $(0, +\infty)$.
- Thus, X is neither discrete nor continuous.
- It is **mixed**.

3.2 Distribution of a random vector

3.2.1 Joint distribution and marginal distributions

DEFINITION 3.3

If X and Y are random variables, then the pair (X, Y) is a **random vector**. Its distribution is called the **joint distribution** of X and Y. Individual distributions of X and Y are then called the **marginal distributions**.

- Although we talk about two random variables in this section,
  all the concepts extend to
  a vector $(X_1, X_2, \ldots, X_n)$
  of n components and its joint distribution.
- Similarly to a single variable,
  the joint distribution of a vector is
  a collection of probabilities
  for a vector (X, Y) to take a value (x, y).
- Recall that two vectors are equal,
  $(X, Y) = (x, y)$, if $X = x$ and $Y = y$.
- This "and" means the intersection, therefore,
  the joint probability mass function of X and Y is
  $P(x, y) = P\{(X, Y) = (x, y)\} = P\{X = x \cap Y = y\}$.
- Again,
  $\{(X, Y) = (x, y)\}$ are exhaustive and mutually exclusive events
  for different pairs (x, y), therefore,
  $$\sum_x \sum_y P(x, y) = 1$$
- The joint distribution of (X, Y) carries the complete information
  about the behavior of this random vector.
- In particular,
  the marginal probability mass functions of X and Y can be obtained from the joint pmf by the Addition Rule.

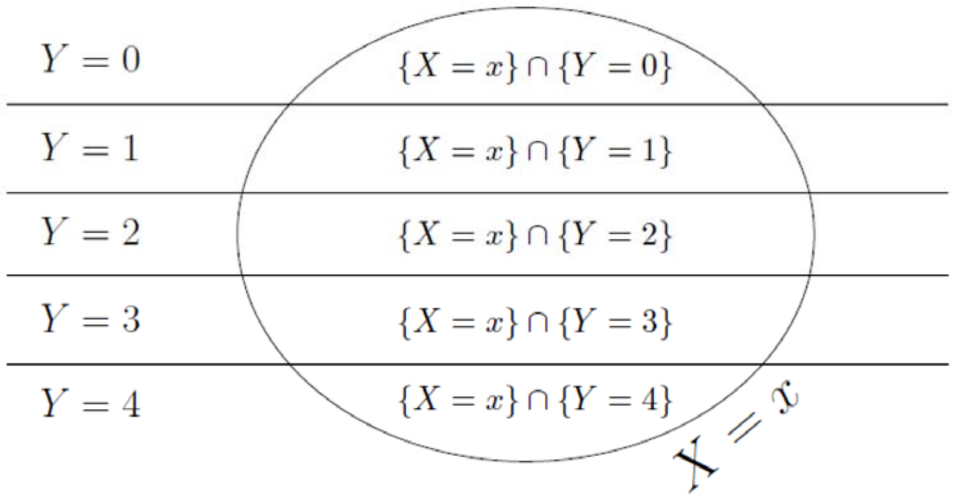


FIGURE 3.2: *Addition Rule: computing marginal probabilities from the joint distribution.*

Addition Rule

$$\boxed{\begin{array}{rclcl} P_X(x) & = & \boldsymbol{P}\{X = x\} & = & \displaystyle\sum_y P_{(X,Y)}(x, y) \\ P_Y(y) & = & \boldsymbol{P}\{Y = y\} & = & \displaystyle\sum_x P_{(X,Y)}(x, y) \end{array}}$$

3.2.2 Independence of random variables

DEFINITION 3.4

Random variables X and Y are **independent** if

$$P_{(X,Y)}(x, y) = P_X(x)P_Y(y)$$

for *all* values of x and y. This means, events $\{X = x\}$ and $\{Y = y\}$ are independent for all x and y; in other words, variables X and Y take their values independently of each other.

Example 3.6

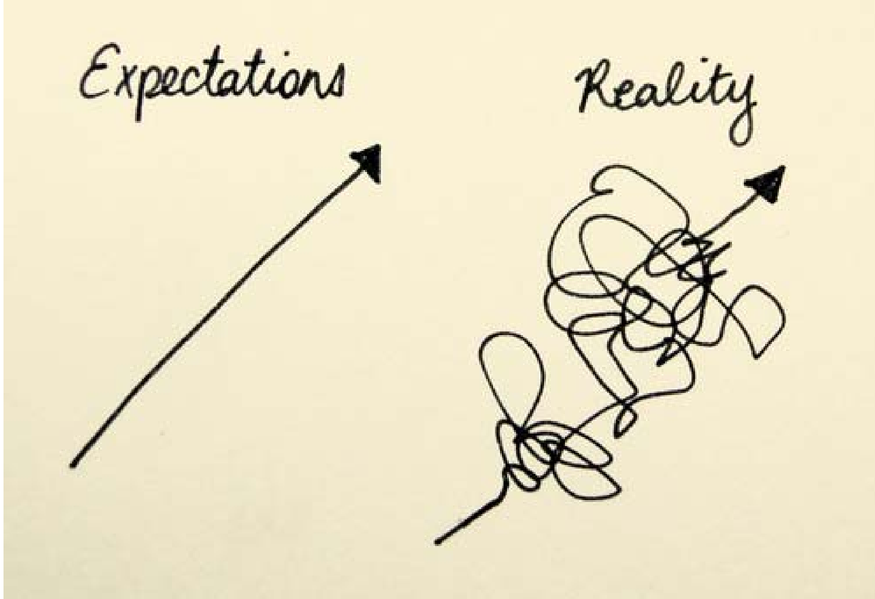
- A program consists of two modules.
- The number of errors, X, in the first module and the number of errors, Y, in the second module have the joint distribution,

  $P(0, 0) = P(0, 1) = P(1, 0) = 0.2,$
  $P(1, 1) = P(1, 2) = P(1, 3) = 0.1,$
  $P(0, 2) = P(0, 3) = 0.05.$
- Find (a) the marginal distributions of X and Y,
  (b) the probability of no errors in the first module,
  and (c) the distribution of the total number of errors in the program.
- Also, (d) find out if errors in the two modules occur independently.
- Solution.
  - It is convenient to organize the joint pmf of X and Y in a table.
  - Adding rowwise and columnwise, we get the marginal pmfs,

$P_{(X,Y)}(x,y)$		y = 0	1	2	3	$P_X(x)$
x	0	0.20	0.20	0.05	0.05	0.50
	1	0.20	0.10	0.10	0.10	0.50
$P_Y(y)$		0.40	0.30	0.15	0.15	1.00

  - This solves (a).

- (b) $P_X(0) = 0.50$.
- (c) Let $Z = X + Y$ be the total number of errors.
- To find the distribution of Z, we first identify its possible values, then find the probability of each value.
- We see that Z can be as small as 0 and as large as 4.
- Then,
  $P_Z(0) = \boldsymbol{P}\{X + Y = 0\}$
  $= \boldsymbol{P}\{X = 0 \cap Y = 0\} = P(0, 0) = 0.20,$
  $P_Z(1) = \boldsymbol{P}\{X = 0 \cap Y = 1\} + \boldsymbol{P}\{X = 1 \cap Y = 0\}$
  $= P(0, 1) + P(1, 0) = 0.20 + 0.20 = 0.40,$
  $P_Z(2) = P(0, 2) + P(1, 1) = 0.05 + 0.10 = 0.15,$
  $P_Z(3) = P(0, 3) + P(1, 2) = 0.05 + 0.10 = 0.15,$
  $P_Z(4) = P(1, 3) = 0.10.$
- It is a good check to verify that $\sum_z P(z) = 1$
- (d) To decide on the independence of X and Y, check if their joint pmf factors into a product of marginal pmfs.
- We see that $P(X,Y)(0, 0) = 0.2$ indeed equals $P_X(0)P_Y(0) = (0.5)(0.4)$.
- Keep checking...
- Next, $P(X,Y)(0, 1) = 0.2$ whereas $P_X(0)P_Y(1) = (0.5)(0.3) = 0.15$.
- There is no need to check further.
- We found a pair of x and y that violates the formula for independent random variables.
- Therefore, the numbers of errors in two modules are dependent.



3.3 Expectation and variance

3.3.1 Expectation

DEFINITION 3.5

Expectation or **expected value** of a random variable X is its mean, the average value.

Example 3.7

- Consider a variable that takes values 0 and 1 with probabilities $P(0) = \mathrm{P}(1) = 0.5$.
- That is,

$$X = \begin{cases} 0 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases}$$

- Observing this variable many times, we shall see $X = 0$ about 50% of times and $X = 1$ about 50% of times.
- The average value of X will then be close to 0.5, so it is reasonable to have $E(X) = 0.5$.

Example 3.8

- Suppose that $P(0) = 0.75$ and $P(1) = 0.25$.
- Then, in a long run, X is equal 1 only 1/4 of times, otherwise it equals 0.
- Suppose we earn \$1 every time we see $X = 1$.
- On the average, we earn \$1 every four times, or \$0.25 per each observation.
- Therefore, in this case $E(X) = 0.25$.

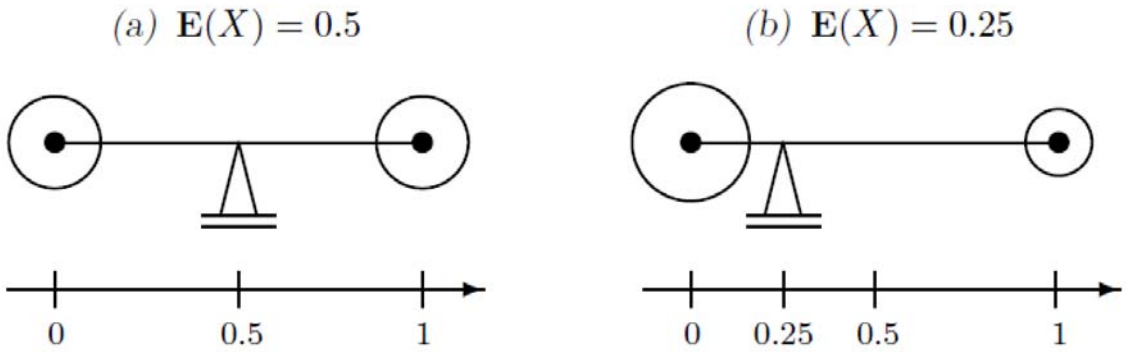


FIGURE 3.3: *Expectation as a center of gravity.*

Expectation, discrete case

$$\boxed{\mu = \mathbf{E}(X) = \sum_x x P(x)}$$

3.3.2 Expectation of a function

- Often we are interested in another variable, Y, that is a function of X.
- For example,
  downloading time depends on the connection speed,
  profit of a computer store depends on the number of computers sold,
  and bonus of its manager depends on this profit.
- Expectation of $Y = g(X)$ is computed by a similar formula,

$$E\{g(X)\} = \sum_x g(x)P(x) \tag{3.4}$$

3.3.3 Properties

Properties of expectations

$$\mathbf{E}(aX + bY + c) = a\,\mathbf{E}(X) + b\,\mathbf{E}(Y) + c$$

In particular,

$$\begin{aligned} \mathbf{E}(X+Y) &= \mathbf{E}(X) + \mathbf{E}(Y) \\ \mathbf{E}(aX) &= a\,\mathbf{E}(X) \\ \mathbf{E}(c) &= c \end{aligned} \qquad (3.5)$$

For **independent** X and Y,

$$\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$$

Proof:

$$
\begin{aligned}
\mathbf{E}(aX+bY+c) &= \sum_x \sum_y (ax+by+c) P_{(X,Y)}(x,y) \\
&= \sum_x ax \sum_y P_{(X,Y)}(x,y) + \sum_y by \sum_x P_{(X,Y)}(x,y) + c \sum_x \sum_y P_{(X,Y)}(x,y) \\
&= a \sum_x x P_X(x) + b \sum_y y P_Y(y) + c.
\end{aligned}
$$

$$
\mathbf{E}(XY) = \sum_x \sum_y (xy) P_X(x) P_Y(y) = \sum_x x P_X(x) \sum_y y P_Y(y) = \mathbf{E}(X)\,\mathbf{E}(Y).
$$

Example 3.9. In Example 3.6 on p. 46,

$$\begin{aligned}
\mathbf{E}(X) &= (0)(0.5) + (1)(0.5) = 0.5 \text{ and} \\
\mathbf{E}(Y) &= (0)(0.4) + (1)(0.3) + (2)(0.15) + (3)(0.15) = 1.05,
\end{aligned}$$

therefore, the expected total number of errors is

$$\mathbf{E}(X+Y) = 0.5 + 1.05 = \quad \mathbf{1.55}$$

$P_{(X,Y)}(x,y)$		y				$P_X(x)$
		0	1	2	3	
x	0	0.20	0.20	0.05	0.05	0.50
	1	0.20	0.10	0.10	0.10	0.50
$P_Y(y)$		0.40	0.30	0.15	0.15	1.00

3.3.4 Variance and standard deviation

- **Example 3.10**
  - Here is a rather artificial but illustrative scenario.
  - Consider two users.
  - One receives either 48 or 52 e-mail messages per day, with a 50-50% chance of each.
  - The other receives either 0 or 100 e-mails, also with a 50-50% chance.
  - What is a common feature of these two distributions, and how are they different?

- We see that both users receive the same average number of e-mails:
  $E(X) = E(Y) = 50$.
- However, in the first case, the actual number of e-mails is always close to 50, whereas it always differs from it by 50 in the second case.
- The first random variable, X, is more stable; it has low variability.
- The second variable, Y , has high variability.

Variance

$$\mu = \mathbf{E}(X)$$

DEFINITION 3.6

Variance of a random variable is defined as the expected squared deviation from the mean. For discrete random variables, variance is

$$\sigma^2 = \mathrm{Var}(X) = \mathbf{E}\,(X - \mathbf{E}X)^2 = \sum_x (x-\mu)^2 P(x)$$

$$\mathrm{Var}(X) = \mathbf{E}(X^2) - \mu^2.$$

Standard deviation

DEFINITION 3.7

Standard deviation is a square root of variance,

$$\sigma = \mathrm{Std}(X) = \sqrt{\mathrm{Var}(X)}$$

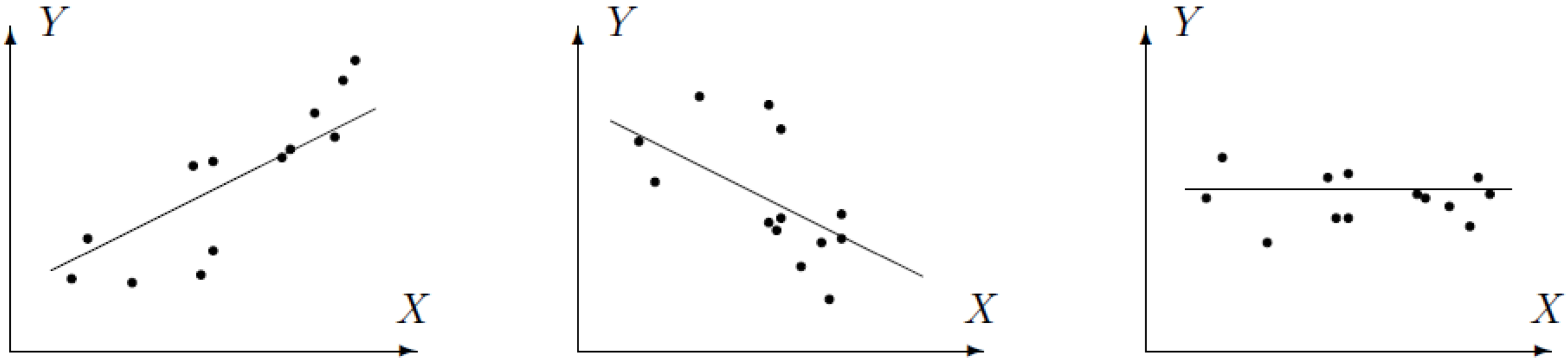
3.3.5 Covariance and correlation

DEFINITION 3.8

Covariance $\sigma_{XY} = \mathrm{Cov}(X, Y)$ is defined as

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \mathbf{E}\{(X - \mathbf{E}X)(Y - \mathbf{E}Y)\} \\ &= \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y) \end{aligned}$$

It summarizes interrelation of two random variables.



(a) $\mathrm{Cov}(X, Y) > 0$ (b) $\mathrm{Cov}(X, Y) < 0$ (c) $\mathrm{Cov}(X, Y) = 0$

FIGURE 3.4: *Positive, negative, and zero covariance.*

DEFINITION 3.9

Correlation coefficient between variables X and Y is defined as

$$\rho = \frac{\mathrm{Cov}(X,Y)}{(\,\mathrm{Std}X)(\,\mathrm{Std}Y)}$$

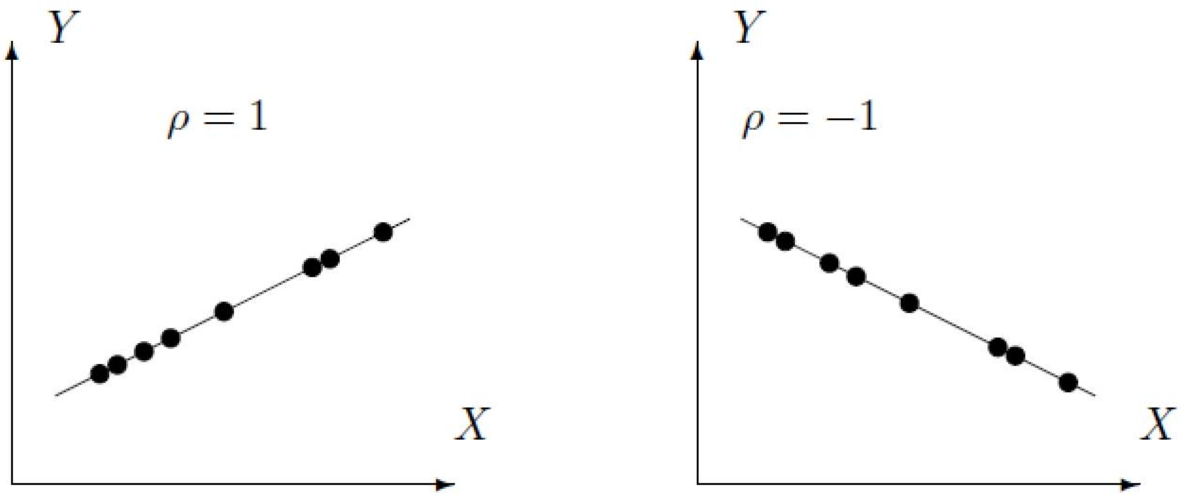


FIGURE 3.5: *Perfect correlation:* $\rho = \pm 1$.

3.3.6 Properties

Properties of variances and covariances

$$\mathrm{Var}(aX + bY + c) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y)$$

$$\begin{aligned}&\mathrm{Cov}(aX + bY, cZ + dW)\\ &\quad = ac\,\mathrm{Cov}(X,Z) + ad\,\mathrm{Cov}(X,W) + bc\,\mathrm{Cov}(Y,Z) + bd\,\mathrm{Cov}(Y,W)\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{Cov}(Y,X)\\ \rho(X,Y) &= \rho(Y,X)\end{aligned} \tag{3.7}$$

In particular,

$$\begin{aligned}\mathrm{Var}(aX + b) &= a^2\,\mathrm{Var}(X)\\ \mathrm{Cov}(aX + b, cY + d) &= ac\,\mathrm{Cov}(X,Y)\\ \rho(aX + b, cY + d) &= \rho(X,Y)\end{aligned}$$

For independent X and Y,

$$\begin{aligned}\mathrm{Cov}(X,Y) &= 0\\ \mathrm{Var}(X + Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y)\end{aligned}$$

$$
\begin{aligned}
\operatorname{Var}(aX+bY+c) &= \mathbf{E}\{aX+bY+c-\mathbf{E}(aX+bY+c)\}^2 \\
&= \mathbf{E}\{(aX-a\,\mathbf{E}X)+(bY-b\,\mathbf{E}Y)+(c-c)\}^2 \\
&= \mathbf{E}\{a(X-\mathbf{E}X)\}^2+\mathbf{E}\{b(Y-\mathbf{E}Y)\}^2+\mathbf{E}\{a(X-\mathbf{E}X)b(Y-\mathbf{E}Y)\} \\
&\quad +\mathbf{E}\{b(Y-\mathbf{E}Y)a(X-\mathbf{E}X)\} \\
&= a^2\operatorname{Var}(X)+b^2\operatorname{Var}(Y)+2ab\operatorname{Cov}(X,Y).
\end{aligned}
$$

Example 3.11. Continuing Example 3.6, we compute

x	$P_X(x)$	$xP_X(x)$	$x - \mathbf{E}X$	$(x - \mathbf{E}X)^2 P_X(x)$
0	0.5	0	-0.5	0.125
1	0.5	0.5	0.5	0.125
$\mu_X = 0.5$			$\sigma_X^2 = 0.25$	

and (using the second method of computing variances)

y	$P_Y(y)$	$yP_Y(y)$	y^2	$y^2 P_Y(y)$
0	0.4	0	0	0
1	0.3	0.3	1	0.3
2	0.15	0.3	4	0.6
3	0.15	0.45	9	1.35
$\mu_Y = 1.05$			$\mathbf{E}(Y^2) = 2.25$	

<u>Result</u>: $\text{Var}(X) = 0.25$, $\text{Var}(Y) = 2.25 - 1.05^2 = 1.1475$, $\text{Std}(X) = \sqrt{0.25} = 0.5$, and $\text{Std}(Y) = \sqrt{1.1475} = 1.0712$.

Also,

$$\mathbf{E}(XY) = \sum_x \sum_y xyP(x,y) = (1)(1)(0.1) + (1)(2)(0.1) + (1)(3)(0.1) = 0.6$$

(the other five terms in this sum are 0). Therefore,

$$\text{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y) = 0.6 - (0.5)(1.05) = 0.075$$

and

$$\rho = \frac{\text{Cov}(X,Y)}{(\text{Std}X)(\text{Std}Y)} = \frac{0.075}{(0.5)(1.0712)} = 0.1400.$$

$P_{(X,Y)}(x,y)$		y				$P_X(x)$
		0	1	2	3	
x	0	0.20	0.20	0.05	0.05	0.50
	1	0.20	0.10	0.10	0.10	0.50
$P_Y(y)$		0.40	0.30	0.15	0.15	1.00

<u>NOTATION</u>

$$
\begin{array}{rcl}
\mu \text{ or } \mathbf{E}(X) & = & \text{expectation} \\
\sigma_X^2 \text{ or } \mathrm{Var}(X) & = & \text{variance} \\
\sigma_X \text{ or } \mathrm{Std}(X) & = & \text{standard deviation} \\
\sigma_{XY} \text{ or } \mathrm{Cov}(X,Y) & = & \text{covariance} \\
\rho_{XY} & = & \text{correlation coefficient}
\end{array}
$$

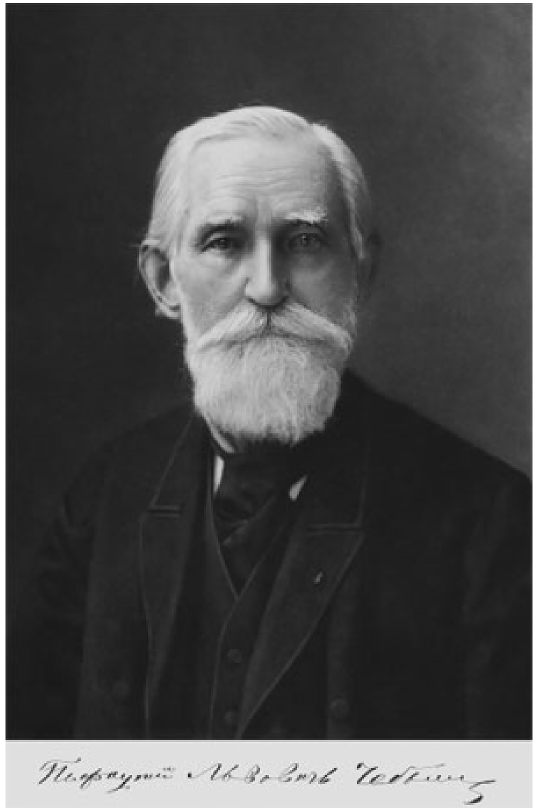
3.3.7 Chebyshev's inequality

Chebyshev's inequality

$$P\{|X-\mu|>\varepsilon\} \le \left(\frac{\sigma}{\varepsilon}\right)^2 \qquad (3.8)$$

for any distribution with expectation μ and variance σ^2 and for any positive ε.

$$\begin{aligned}\sigma^2 &= \sum_{\text{all } x}(x-\mu)^2 P(x) \ge \sum_{\text{only } x:|x-\mu|>\varepsilon}(x-\mu)^2 P(x)\\ &\ge \sum_{x:|x-\mu|>\varepsilon}\varepsilon^2 P(x) = \varepsilon^2 \sum_{x:|x-\mu|>\varepsilon} P(x) = \varepsilon^2 P\{|x-\mu|>\varepsilon\}.\end{aligned}$$



Example 3.12

- Suppose the number of errors in a new software has expectation $\mu = 20$ and a standard deviation of 2.
- According to (3.8), there are more than 30 errors with probability

$$\boldsymbol{P}\{X > 30\} \leq \boldsymbol{P}\{|X - 20| > 10\} \leq \left(\frac{2}{10}\right)^2 = 0.04.$$

- However, if the standard deviation is 5 instead of 2, then the probability of more than 30 errors can only be bounded by

$$\left(\frac{5}{10}\right)^2 = 0.25.$$

3.3.8 Application to finance

- Example 3.13 (Construction of an optimal portfolio).
  - We would like to invest \$10,000 into shares of companies XX and YY.
  - Shares of XX cost \$20 per share.
  - The market analysis shows that their expected return is \$1 per share with a standard deviation of \$0.5.
  - Shares of YY cost \$50 per share, with an expected return of \$2.50 and a standard deviation of \$1 per share, and returns from the two companies are independent.
  - In order to maximize the expected return and minimize the risk (standard deviation or variance), is it better to invest
    (A) all \$10,000 into XX,
    (B) all \$10,000 into YY,
    or (C) \$5,000 in each company?

- Solution.
  - Let X be the actual (random) return from each share of XX, and Y be the actual return from each share of YY.
  - Compute the expectation and variance of the return for each of the proposed portfolios (A, B, and C).

a. At \$20 a piece, we can use \$10,000 to buy 500 shares of XX collecting a profit of $A = 500X$.
   Using **(3.5)** and **(3.7)**,
   $\mathbf{E}(A) = 500\mathbf{E}(X) = (500)(1) = 500$;
   $\mathrm{Var}(A) = 500^2\mathrm{Var}(X) = 500^2(0.5)^2 = 62{,}500$.

b. Investing all \$10,000 into YY,
we buy 10,000/50=200 shares of it
and collect a profit of $B = 200Y$,
$\mathbf{E}(B) = 200\mathbf{E}(Y) = (200)(2.50) = 500$;
$\mathrm{Var}(A) = 200^2\mathrm{Var}(Y) = 200^2(1)^2 = 40{,}000$.

c. Investing \$5,000 into each company
makes a portfolio consisting
of 250 shares of XX
and 100 shares of YY;
the profit in this case will be
$C = 250X + 100Y$.
Following **(3.7)** for independent X and Y,
$\mathbf{E}(C) = 250\mathbf{E}(X) + 100\mathbf{E}(Y)$
$= 250 + 250 = 500$;
$\mathrm{Var}(C) = 250^2\mathrm{Var}(X) + 100^2\mathrm{Var}(Y)$
$= 250^2(0.5)^2 + 100^2(1)^2 = 25{,}625$.

- Result:
  - The expected return is the same for each of the proposed three portfolios because each share of each company is expected to return 1/20 or 2.50/50, which is 5%.
  - In terms of the expected return, all three portfolios are **equivalent**.
  - Portfolio C, where investment is split between two companies, has the lowest variance; therefore, it is the least risky.
  - This supports one of the basic principles in finance: to **minimize the risk, diversify the portfolio**.

Example 3.15 (Optimizing even further)

- So, after all, with \$10,000 to invest, what is the most optimal portfolio consisting of shares of XX and YY, given their correlation coefficient of $\rho = -0.2$?
- This is an **optimization problem**.
- Suppose t dollars are invested into XX and $(10{,}000 - t)$ dollars into YY, with the resulting profit is C_t.
- This amounts for $t/20$ shares of X and $(10{,}000 - t)/50 = 200 - t/50$ shares of YY.
- Plans A and B correspond to $t = 10{,}000$ and $t = 0$.
- The expected return remains constant at \$500.
- If $\mathrm{Cov}(X, Y) = -0.1$, as in **Example 3.14** above, then
  $\mathrm{Var}(C_t) = \mathrm{Var}\{tX + (200 - t/50)Y\}$
  $= (t/20)^2\mathrm{Var}(X) + (200 - t/50)^2\mathrm{Var}(Y) + 2(t/20)(200 - t/50)\mathrm{Cov}(X, Y)$
  $= (t/20)^2(0.5)^2 + (200 - t/50)^2(1)^2 + 2(t/20)(200 - t/50)(-0.1)$
  $= \dfrac{49t^2}{40{,}000} - 10t + 40{,}000.$
- See the graph of this function in **Figure 3.6**.
- Minimum of this variance is found at $t^* = 10/(\frac{49t^2}{40{,}000}) = 4081.63$.
- Thus, for the most optimal portfolio, we should invest \$4081.63 into XX and the remaining \$5919.37 into YY.
- Then we achieve the smallest possible risk (variance) of $(\$^2)$19, 592, measured, as we know, in squared dollars.

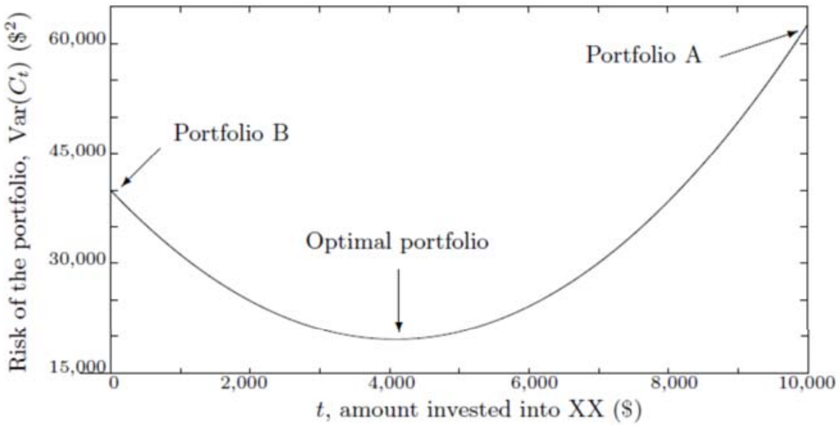
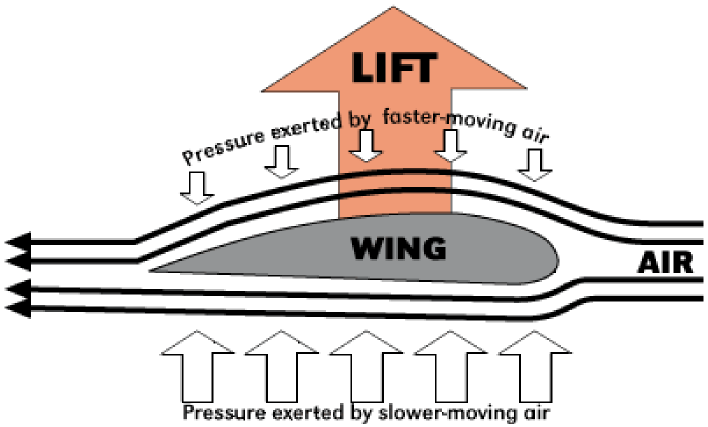


FIGURE 3.6: *Variance of a diversified portfolio.*

3.4 Families of discrete distributions

Bernoulli

SOLO



BERNOULLI FAMILY

SOLO HERMELIN

http://www.solohermelin.com

The Bernoulli family

Nicolaus
1623-1708

Jacob 1654-1705 | Nicolaus 1662-1716 | **Johann** 1667-1748

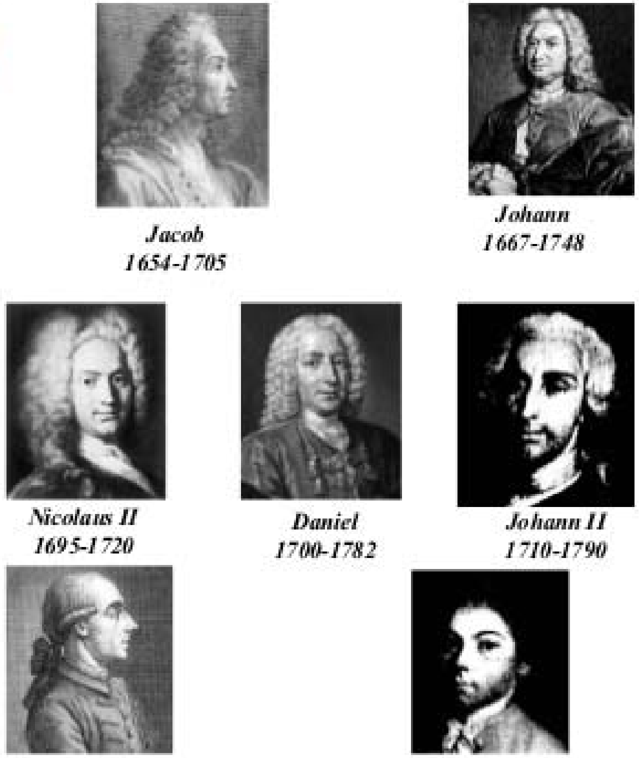
Nicolaus (I) 1687-1759

Nicolaus (II) 1695-1726 | **Daniel** 1700-1782 | **Johann (II)** 1710-1790

Johann (III) 1744-1807 | Daniel (II) 1751-1834 | **Jacob (II)** 1759-1789

Those shown in **bold** above are in our archive

Run This



Jacob 1654-1705

Johann 1667-1748

Nicolaus II 1695-1720

Daniel 1700-1782

Johann II 1710-1790

Johann III 1744-1807

Jacob II 1759-1789

3.4.1 Bernoulli distribution

- DEFINITION 3.10
  - A random variable with two possible values, 0 and 1,
    is called a **Bernoulli variable**,
    its distribution is **Bernoulli distribution**,
    and any experiment with a binary outcome is called a **Bernoulli trial**.

If $P(1) = p$ is the probability of a *success*, then $P(0) = q = 1 - p$ is the probability of a *failure*. We can then compute the expectation and variance as

$$\begin{aligned}
\mathbf{E}(X) &= \sum_x P(x) = (0)(1-p) + (1)(p) = p, \\
\mathrm{Var}(X) &= \sum_x (x-p)^2 P(x) = (0-p)^2(1-p) + (1-p)^2 p \\
&= p(1-p)(p+1-p) = p(1-p).
\end{aligned}$$

Bernoulli distribution

$$\begin{aligned}
p &= \text{probability of success} \\
P(x) &= \begin{cases} q = 1-p & \text{if} \quad x = 0 \\ p & \text{if} \quad x = 1 \end{cases} \\
\mathbf{E}(X) &= p \\
\mathrm{Var}(X) &= pq
\end{aligned}$$

3.4.2 Binomial distribution

- DEFINITION 3.11
  - A variable described as the number of successes in a sequence of independent Bernoulli trials has **Binomial distribution**.
  - Its parameters are n, the number of trials, and p, the probability of success.

$$P(x) = \boldsymbol{P}\{X = x\} = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, \ldots, n, \qquad (3.9)$$

Example 3.16

- As part of a business strategy, randomly selected 20% of new internet service subscribers receive a special promotion from the provider.
- A group of 10 neighbors signs for the service.
- What is the probability that at least 4 of them get a special promotion?

Table A2. Binomial distribution

$$F(x) = \boldsymbol{P}\{X \le x\} = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

n	x	p																		
		.050	.100	.150	.200	.250	.300	.350	.400	.450	.500	.550	.600	.650	.700	.750	.800	.850	.900	.950
1	0	.950	.900	.850	.800	.750	.700	.650	.600	.550	.500	.450	.400	.350	.300	.250	.200	.150	.100	.050
2	0	.903	.810	.723	.640	.563	.490	.423	.360	.303	.250	.203	.160	.123	.090	.063	.040	.023	.010	.003
	1	.998	.990	.978	.960	.938	.910	.878	.840	.798	.750	.698	.640	.578	.510	.438	.360	.278	.190	.098

10	0	.599	.349	.197	.107	.056	.028
	1	.914	.736	.544	.376	.244	.149
	2	.988	.930	.820	.678	.526	.383
	3	.999	.987	.950	.879	.776	.650
	4	1.0	.998	.990	.967	.922	.850

- Solution.
  - We need to find the probability $\boldsymbol{P}\{X \ge 4\}$, where X is the number of people, out of 10, who receive a special promotion.
  - This is the number of successes in 10 Bernoulli trials, therefore, X has Binomial distribution with parameters $n = 10$ and $p = 0.2$.
  - From **Table A2**, $\boldsymbol{P}\{X \ge 4\} = 1 - F(3) = 1 - 0.8791 = 0.1209$.

$$\mathbf{E}(X) = \sum_{x=0}^{n} x \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^{n-x} = \ldots ?$$

$$X = X_1 + \ldots + X_n.$$

$$\mathbf{E}(X) = \mathbf{E}(X_1 + \ldots + X_n) = \mathbf{E}(X_1) + \ldots + \mathbf{E}(X_n) = p + \ldots + p = np$$

and

$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = npq.$$

Binomial distribution

$$\begin{array}{lcl} n & = & \text{number of trials} \\ p & = & \text{probability of success} \\ P(x) & = & \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^{n-x} \\ \mathbf{E}(X) & = & np \\ \mathrm{Var}(X) & = & npq \end{array}$$

Example 3.17

- An exciting computer game is released.
- 60 % of players complete all the levels.
- 30 % of them will then buy an advanced version of the game.
- Among 15 users,
  what is the expected number of people who will buy the advanced version?
- What is the probability
  that at least two people will buy it?

- Solution.
  - Let X be the number of people (successes), among the mentioned 15 users (trials), who will buy the advanced version of the game.
  - It has Binomial distribution with $n = 15$ trials and the probability of success
    $$p = \boldsymbol{P}\{\text{buy advanced}\} = \boldsymbol{P}\{\text{buy advanced} \mid \text{complete all levels}\}\boldsymbol{P}\{\text{complete all levels}\} = (0.30)(0.60) = 0.18.$$
  - Then we have
    $$E(X) = np = (15)(0.18) = 2.7$$
    and
    $$\boldsymbol{P}\{X \geq 2\} = 1 - P(0) - P(1) = 1 - (1-p)^n - np(1-p)^{n-1} = 0.7813.$$
  - The last probability was computed directly by formula **(3.9)** because the probability of success, 0.18, is not in **Table A2**.

3.4.3 Geometric distribution

- Geometric Series

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 \cdots, \qquad |x| < 1$$

Geometric distribution

DEFINITION 3.12

The number of Bernoulli trials needed to get the first success has **Geometric distribution**.

Example 3.18. A search engine goes through a list of sites looking for a given key phrase. Suppose the search terminates as soon as the key phrase is found. The number of sites visited is Geometric. $\diamond$

Example 3.19. A hiring manager interviews candidates, one by one, to fill a vacancy. The number of candidates interviewed until one candidate receives an offer has Geometric distribution. $\diamond$

- Geometric random variables can take any integer value from 1 to infinity, because one needs at least 1 trial to have the first success, and the number of trials needed is not limited by any specific number.
- (For example, there is no guarantee that among the first 10 coin tosses there will be at least one head.)
- The only parameter is p, the probability of a "success."
- Geometric probability mass function has the form
  $P(\mathrm{x}) = \boldsymbol{P}\{\text{the 1st success occurs on the } x\text{-th trial}\}$
  $= (1-p)^{x-1}p,\ x = 1, 2, \ldots,$
  which is the probability of $(x-1)$ failures followed by one success.
- Comparing with **(3.9)**, there is no number of combinations in this formula because only one outcome has the first success coming on the x-th trial.
- This is the first time we see an **unbounded** random variable, that is, with no upper bound.

- Variable X can take any positive integer value, from 1 to ∞.
- It is insightful to check whether $\sum_x P(x) = 1$, as it should hold for all pmf.
- Indeed,
  $$\sum_x P(x) = \sum_{x=1}^{\infty}(1-p)^{x-1}p = \frac{(1-p)^0}{1-(1-p)}p = 1$$
  where we noticed that the sum on the left is a **geometric series** that gave its name to Geometric distribution.
- Finally, Geometric distribution has expectation $\mu = 1/p$ and variance $\sigma^2 = (1-p)/p^2$.
- Proof:
  - The geometric series $s(q) = \sum_0^\infty q^x$ equals $(1-q)^{-1}$.
  - Taking derivatives with respect to q,
    $\frac{1}{(1-q)^2} = \left(\frac{1}{1-q}\right)' = s'(q) = (\sum_0^\infty q^x)' = \sum_0^\infty xq^{x-1} = \sum_1^\infty xq^{x-1}$
  - It follows that for a Geometric variable X with parameter $p = 1-q$,
    $E(X) = \sum_1^\infty xq^{x-1} = \frac{1}{(1-q)^2}p = \frac{1}{p}$.
  - Derivation of the variance is similar.
  - One has to take the second derivative of $s(q)$ and get, after a few manipulations, an expression for $\sum x^2q^{x-1}$

Geometric distribution

$$\begin{aligned} p &= \text{probability of success} \\ P(x) &= (1-p)^{x-1}p, \quad x = 1, 2, \ldots \\ \mathbf{E}(X) &= \frac{1}{p} \\ \mathrm{Var}(X) &= \frac{1-p}{p^2} \end{aligned}$$

Example 3.20 (St. Petersburg Paradox)

- This paradox was noticed
  by a Swiss mathematician Daniel Bernoulli (1700–1782),
  a nephew of Jacob.
- It describes a gambling strategy
  that enables one to win any desired amount of money
  with probability one.
- Isn't it a very attractive strategy?
- It's real, there is no cheating!
- Consider a game that can be played any number of times.
- Rounds are independent,
  and each time your winning probability is p.
- The game does not have to be favorable to you or even fair;
  this p can be any positive probability.
- For each round, you bet some amount x.
- In case of a success, you win x.
- If you lose a round, you lose x.
- The strategy is simple.
- Your initial bet is the amount that you desire to win eventually.
- Then, if you win a round, stop.
- If you lose a round, double your bet and continue.
- Let the desired profit be \$100.
- The game will progress as follows.
- Sooner or later, the game will stop, and at this moment,
  your balance will be \$100. Guaranteed!
- However, this is not what D. Bernoulli called a paradox.
- How many rounds should be played?
- Since each round is a Bernoulli trial,
  the number of them, X,
  until the first win is a Geometric random variable with parameter p.

Round	Bet	Balance...	
		... if lose	... if win
1	100	-100	$+100$ and stop
2	200	-300	$+100$ and stop
3	400	-700	$+100$ and stop
...	...	...	...

- Is the game endless?
- No, on the average, it will last $E(X) = 1/p$ rounds.
- In a fair game with $p = 1/2$,
  one will need 2 rounds, on the average, to win the desired amount.
- In an "unfair" game, with $p < 1/2$,
  it will take longer to win,
  but still a finite number of rounds.
- For example, if $p = 0.2$, i.e., one win in five rounds, then on the average, one stops after $1/p = 5$ rounds.
- This is not a paradox yet.
- Finally, how much money does one need to have in order to be able to follow this strategy?
- Let Y be the amount of the last bet.
- According to the strategy, $Y = 100 \cdot 2^{X-1}$.
- It is a discrete random variable whose expectation equals
  $$E(Y) = \sum_x (100 \cdot 2^{x-1})\, P_X(x)$$
  $$= 100 \sum_1^\infty 2^{x-1}(1-p)^{x-1}\, p$$
  $$= 100p \sum_1^\infty (2(1-p))^{x-1} = \begin{cases} \frac{100p}{2(1-p)} & \text{if } p > 1/2 \\ +\infty & \text{otherwise} \end{cases}$$
- This is the St. Petersburg Paradox!
- A random variable that is always finite has an infinite expectation!
- Even when the game is fair offering a 50-50 chance to win, one has to be (onthe average!) infinitely wealthy to follow this strategy.
- To the best of our knowledge,
  every casino has a limit on the maximum bet, making sure gamblers cannot fully apply this St. Petersburg strategy.
- When such a limit is enforced,
  it can be proved that a winning strategy does not exist.

3.4.4 Negative Binomial Distribution

DEFINITION 3.13

In a sequence of independent Bernoulli trials, the number of trials needed to obtain k successes has **Negative Binomial distribution.**

$$
\begin{aligned}
P(x) &= \boldsymbol{P}\{\text{ the } x\text{-th trial results in the } k\text{-th success }\} \\
&= \boldsymbol{P}\left\{\begin{array}{c}(k-1) \text{ successes in the first } (x-1) \text{ trials,} \\ \text{and the last trial is a success}\end{array}\right\} \\
&= \binom{x-1}{k-1}(1-p)^{x-k}p^k.
\end{aligned}
$$

- Negative Binomial distribution has two parameters, k and p.
- With $k = 1$, it becomes Geometric.
- Also, each Negative Binomial variable can be represented as a sum of independent Geometric variables,

$$X = X_1 + \ldots + X_k, \tag{3.11}$$

with the same probability of success p.

$$\begin{aligned}
\mathbf{E}(X) &= \mathbf{E}(X_1 + \ldots + X_k) = \frac{k}{p}; \\
\mathrm{Var}(X) &= \mathrm{Var}(X_1 + \ldots + X_k) = \frac{k(1-p)}{p^2}.
\end{aligned}$$

Negative Binomial distribution

$$
\begin{aligned}
k &= \text{number of successes} \\
p &= \text{probability of success} \\
P(x) &= \binom{x-1}{k-1}(1-p)^{x-k}p^k, \quad x = k, k+1, \ldots \\
\mathbf{E}(X) &= \frac{k}{p} \\
\mathrm{Var}(X) &= \frac{k(1-p)}{p^2}
\end{aligned}
$$

Example 3.21 (Sequential testing)

- In a recent production,
  5% of certain electronic components are defective.
- We need to find 12 non-defective components
  for our 12 new computers.
- Components are tested until 12 non-defective ones are found.
- What is the probability that more than
  15 components will have to be tested?
- Solution.
  - Let X be the number of components tested until 12 non-defective ones are found.
  - It is a number of trials needed to see
    12 successes,
    hence X has Negative Binomial distribution
    with $k = 12$ and $p = 0.05$.

- We need $\boldsymbol{P}\{X > 15\} = = \sum_{16}^{\infty} P(x)$ or $1 - F(15)$;
  however,
  there is no table of Negative Binomial distribution in the Appendix,
  and applying the formula for $P(x)$ directly is rather cumbersome.
- What would be a quick solution?
- Virtually any Negative Binomial problem can be solved by a Binomial distribution.
- Although X is not Binomial at all,
  the probability $\boldsymbol{P}\{X > 15\}$ can be related to
  some Binomial variable.

- In our example,

$$\boldsymbol{P}\{X > 15\} = \boldsymbol{P}\{\text{more than 15 trials needed to get 12 successes}\}$$
$$= \boldsymbol{P}\{\text{15 trials are not sufficient}\}$$
$$= \boldsymbol{P}\{\text{there are fewer than 12 successes in 15 trials}\}$$
$$= \boldsymbol{P}\{Y < 12\},$$

  where Y is the number of successes (non-defective components) in 15 trials, which is a Binomial variable with parameters $n = 15$ and $p = 0.95$.

- From **Table A2** on **p. 412**,

$$\boldsymbol{P}\{X > 15\} = \boldsymbol{P}\{Y < 12\} = \boldsymbol{P}\{Y \leq 11\} = F(11) = 0.0055.$$

- This technique,
  expressing a probability about one random variable
  in terms of another random variable, is rather useful.

- Soon it will help us relate Gamma and Poisson distributions
  and simplify computations significantly.

http://www.di-mgt.com.au/binomial-calculator.html

Binomial distribution calculator

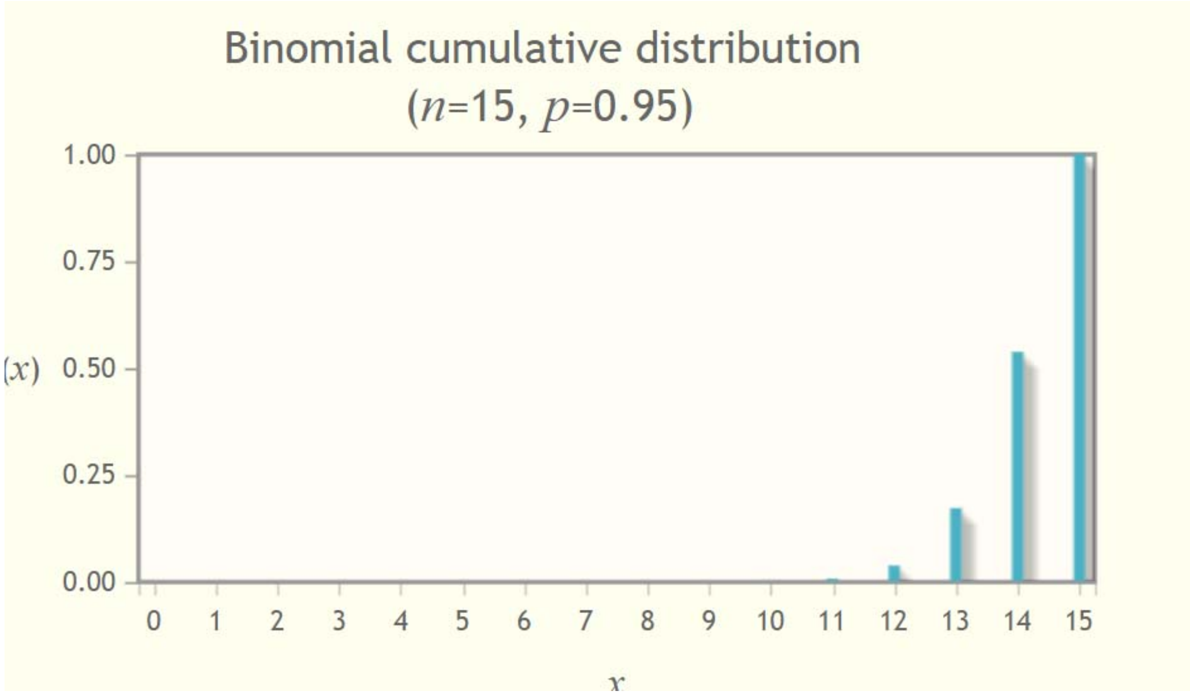
This calculates a table of the binomial distribution for given parameters a
your values of n and p below. There is more on the theory and use of the
click on **Show graph** to see the graphs.

Number of trials, n: 15 (n an integer > 0)

Probability of success, p: 0.95 ($0 < p < 1$)

Calculate table   Show graph

Binomial distribution (n=15, p=0.95)			
	f(x)	F(x)	1 - F(x)
x	Pr[$X = x$]	Pr[$X \le x$]	Pr[$X > x$]
0	0.0000	0.0000	1.0000
1	0.0000	0.0000	1.0000
2	0.0000	0.0000	1.0000
3	0.0000	0.0000	1.0000
4	0.0000	0.0000	1.0000
5	0.0000	0.0000	1.0000
6	0.0000	0.0000	1.0000
7	0.0000	0.0000	1.0000
8	0.0000	0.0000	1.0000
9	0.0000	0.0001	0.9999
10	0.0006	0.0006	0.9994
11	0.0049	0.0055	0.9945
12	0.0307	0.0362	0.9638
13	0.1348	0.1710	0.8290
14	0.3658	0.5367	0.4633
15	0.4633	1.0000	0.0000



3.4.5 Poisson Distribution

DEFINITION 3.14

The number of rare events occurring within a fixed period of time has **Poisson distribution.**

Poisson distribution

$$\begin{aligned} \lambda &= \text{frequency, average number of events} \\ P(x) &= e^{-\lambda}\frac{\lambda^x}{x!},\ x = 0, 1, 2, \ldots \\ \mathbf{E}(X) &= \lambda \\ \mathrm{Var}(X) &= \lambda \end{aligned}$$

PROOF: This probability mass function satisfies $\sum_0^\infty P(x) = 1$ because the Taylor series for $f(\lambda) = e^\lambda$ at $\lambda = 0$ is

$$e^\lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}, \tag{3.13}$$

and this series converges for all λ.

The expectation $\mathbf{E}(X)$ can be derived from (3.13) as

$$\mathbf{E}(X) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = 0 + e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1} \cdot \lambda}{(x-1)!} = e^{-\lambda} \cdot e^{\lambda} \cdot \lambda = \lambda.$$

Similarly,

$$\mathbf{E}(X^2) - \mathbf{E}(X) = \sum_{x=0}^{\infty} (x^2 - x) P(x) = \sum_{x=2}^{\infty} x(x-1) e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2} \cdot \lambda^2}{(x-2)!} = \lambda^2,$$

and therefore,

$$\mathrm{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X) = (\lambda^2 + \mathbf{E}(X)) - \mathbf{E}^2(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

$\square$

FYI: Taylor Series

- https://www.youtube.com/watch?v=3d6DsjIBzJ4

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

Example 3.22 (New accounts)

- Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day.

a. What is the probability that more than 8 new accounts will be initiated today?
   - New account initiations qualify as rare events because no two customers open accounts simultaneously.
   - Then the number X of today's new accounts has Poisson distribution with parameter $\lambda = 10$.
   - From Table A3, $\boldsymbol{P}\{X > 8\} = 1 - F_X(8) = 1 - 0.333 = 0.667$.

b. What is the probability that more than 16 accounts will be initiated within 2 days?
   - The number of accounts, Y, opened within 2 days does not equal $2X$.
   - Rather, Y is another Poisson random variable whose parameter equals 20.
   - Indeed, the parameter is the average number of rare events, which, over the period of two days, doubles the one-day average.
   - Using **Table A3** with $\lambda = 20$, $\boldsymbol{P}\{Y > 16\} = 1 - F_Y(16) = 1 - 0.221 = 0.779$.

https://www.easycalculation.com/statistics/poisson-distribution.php

Poisson Distribution Calculator

English

Calculator   Σ Formula

Free Poisson distribution calculation online. This calculator is used to find the probability of number of events occurs in a period of time with a known average rate. Can be used for calculating or creating new math problems.

Poisson Distribution Calculator

To Calculate Poisson Distribution:

Average rate of success(λ):	10
Poisson Random Variable(x):	8

Calculate

Result:

Poisson Distribution:	0.113
Cumulative Poisson Distribution:	0.333

Code to add this calci to your website

3.4.6 Poisson approximation of binomial distribution

Poisson approximation to Binomial

$$\text{Binomial}(n, p) \approx \text{Poisson}(\lambda)$$

where $n \geq 30$, $p \leq 0.05$, $np = \lambda$

Example 3.23 (New accounts, continued)

- Indeed, the situation in **Example 3.22** can be viewed as a sequence of Bernoulli trials.
- Suppose there are $n = 400{,}000$ potential internet users in the area, and on any specific day, each of them opens a new account with probability $p = 0.000025$.
- We see that the number of new accounts is the number of successes, hence a Binomial model with expectation $E(X) = np = 10$ is possible.
- However, a distribution with such extreme n and p is unlikely to be found in any table, and computing its pmf by hand is tedious.
- Instead, one can use Poisson distribution with the same expectation $\lambda = 10$.

Example 3.24

- 97 % of electronic messages are transmitted with no error.
- What is the probability that out of 200 messages, at least 195 will be transmitted correctly?
- Solution.
  - Let X be the number of correctly transmitted messages.
  - It is the number of successes in 200 Bernoulli trials, thus X is Binomial with $n = 200$ and $p = 0.97$.
  - Poisson approximation cannot be applied to X because p is too large.
  - However, the number of failures Y is also Binomial, with parameters $n = 200$ and $q = 0.03$, and it is approximately Poisson with $\lambda = nq = 6$.
  - From **Table A3**,

$$\boldsymbol{P}\{X \geq 195\} = \mathrm{P}\ \{Y \leq 5\} = F_Y(5) \approx 0.446.$$

Example 3.25 (Birthday problem)

- This continues **Exercise 2.25**.
- Consider a class with $N \geq 10$ students.
- Compute the probability that at least two of them have their birthdays on the same day.
- How many students should be in class in order to have this probability above 0.5?
- Solution.
  - Poisson approximation will be used for the number of shared birthdays among all
    $$n = \binom{N}{2} = \frac{N(N-1)}{2}$$
    pairs of students in this class.
  - In each pair, both students are born on the same day with probability $p = 1/365$.
  - Each pair is a Bernoulli trial because the two birthdays either match or don't match.

- Besides, matches in two different pairs are "nearly" independent.
- Therefore, X, the number of pairs sharing birthdays, is "almost" Binomial.
- For $N \geq 10$, $n \geq 45$ is large, and p is small, thus, we shall use Poisson approximation with $\lambda = np = N(N-1)/730$,
  $\boldsymbol{P}\{$there are two students sharing birthday$\}$
  $$= 1 - \boldsymbol{P}\{\text{no matches}\}$$
  $$= 1 - P\{X = 0\} \approx 1 - e^{-\lambda} \approx 1 - e^{\frac{-N^2}{730}}.$$
- Solving the inequality $1 - e^{\frac{-N^2}{730}} > 0.5$, we obtain $N > \sqrt{730 \ln 2} \approx 22.5$.
- That is, in a class of at least $N = 23$ students, there is a more than 50% chance that at least two students were born on the same day of the year!

If there are n people?

- $\mathrm{P}(B_{22}) = .5243$
- $\mathrm{P}(B_{23}) = .4927$

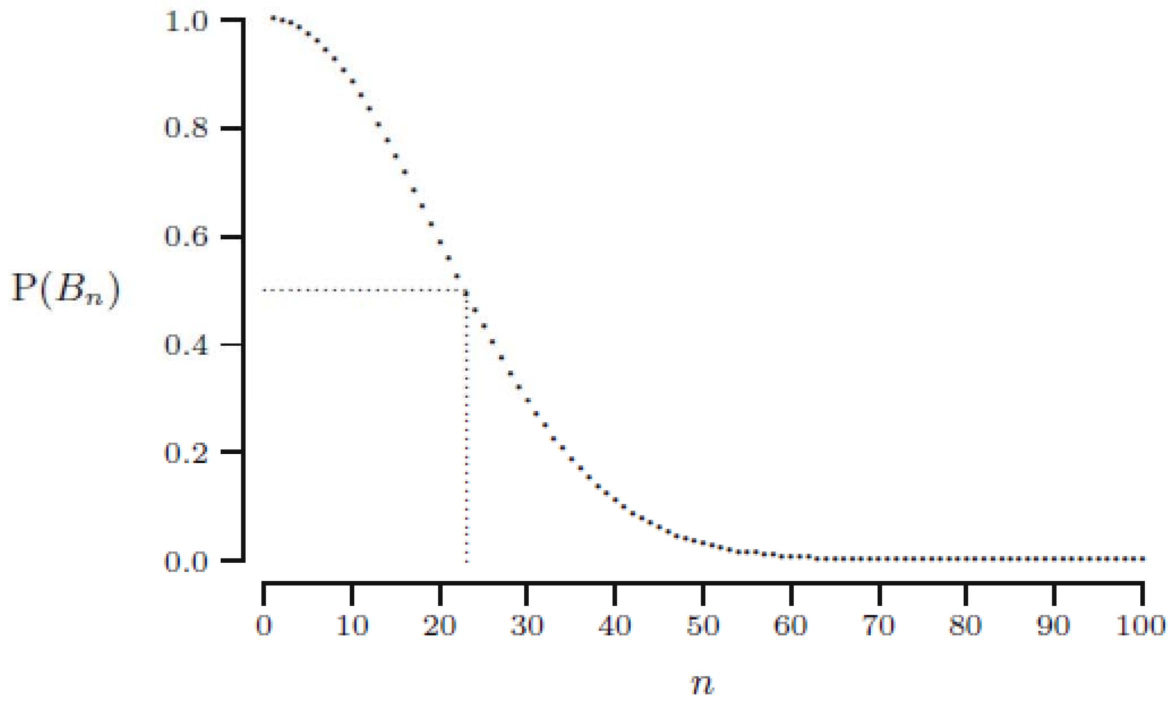


Fig. 3.1. The probability $\mathrm{P}(B_n)$ of no coincident birthdays for $n = 1, \ldots, 100$.

Q&A