Set 1 (a) You are flipping a coin. What is the sample space Ω?

(b) You are interested in the reliability (=the probability of functioning) of a certain (electronic, mechanical, etc.) component. What is the sample space Ω? (You don't need to know the meaning of 'probability' to solve this problem.)

(c) You are flipping three coins. What is the sample space Ω?

(d) You are interested in the reliability (=the probability of functioning) of a wind farm consisting of three wind turbines. What is the sample space Ω? (You don't need to know the meaning of 'probability' to solve this problem.)

Set 2 (a) You are flipping two coins. How many events do we have in total? Why?

(b) What is the difference between ω and ω^{act}?

(c) Let A be an event. When does the event A occur?

(d) You are interested in the reliability of a wind farm consisting of three wind turbines. Your house is fully functional – let's call it event A – when at least two wind turbines are functioning. Express the event A in terms of ω's.

Set 3 (a) In a class of 230 students, there are 200 engineering students (event E), 10 business students (event B), and 20 science students (event S). Assume that all students are viewed equally, and let none of them belong to more than one faculty, that is, let the three events E, B, and S be disjoint. What is the probability of selecting a science or business student?

(b) With the help of the addition rule (what is it?), prove that the equation $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ holds for any two events A and B.

(c) A batch of 150 components were tested in a harsh environment. Some components passed the test but some failed due to high temperature or increased stress. The following failure-frequency table has been reported:

	Stress		
Temperature	PASS	FAIL	
PASS	40	20	60
FAIL	30	60	90
	70	80	150

Calculate the probability of failure due to temperature <u>and</u> stress, and then the probability of failure due to temperature <u>or</u> stress. (To solve this problem correctly, it is useful to specify – and to also visualize by drawing sets and subsets – the sample space and the events of interest.)

(d) Two events A and B are called independent if the probability $\mathbb{P}(A \cap B)$ is equal to the product of $\mathbb{P}(A)$ and $\mathbb{P}(B)$. Show that if the (non-empty) events A and B are disjoint, that is, if $A \cap B = \emptyset$, then the events must be dependent.

Set 4 Use block diagrams (for electrical circuits) to convince yourselves that the following equations make sense and are even 'obvious' (don't forget that W^c means 'failure' F):

(a) de Morgan's rule: $(W_1 \cap W_2)^c = W_1^c \cup W_2^c$ (hint: think of a series system with two components).

(b) another de Morgan's rule: $(W_1 \cup W_2)^c = W_1^c \cap W_2^c$ (hint: think of a parallel system with two components).

(c) the distributive rule: $W_1 \cap (W_2 \cup W_3) = (W_1 \cap W_2) \cup (W_1 \cap W_3)$ (hint: think of a mixed system with one sequential and two parallel components).

Set 5 A wind farm consists of three wind turbines that function independently and each has the same reliability (=the probability of functioning) of 0.99, that is, the reliability is 99% in the 'percentile' layman's language.

(a) Suppose that the three wind turbines are connected into a series system, that is, $W = W_1 \cap W_2 \cap W_3$. Draw the corresponding block diagram and calculate the system's reliability.

(b) Suppose that the three wind turbines are connected into a parallel system, that is, $W = W_1 \cup W_2 \cup W_3$. Draw the corresponding block diagram and calculate the system's reliability. (Warning: 'independent' and 'disjoint' are two different notions, like apples and oranges, and thus can't be interchanged, equated, compared, etc.) Hint: see if the related equation $F = F_1 \cap F_2 \cap F_3$ helps to solve this problem; don't forget that 'failure' F is the complement W^c of 'working' W.

(c) Suppose that the three wind turbines are connected into a mixed system, that is, $W = W_1 \cap (W_2 \cup W_3)$. Draw the corresponding block diagram and calculate the system's reliability.

(d) Which of the above three configurations gives the highest system's reliability? Does it make sense?

The above problems are based on Chapter 2. We are now going into Chapter 3.

Set 6 (a) What is random variable?

(b) Is 'random variable' random?

(c) Is 'random variable' variable?

(d) Let T denote the 'failure time.' Specifically, let T be the mapping from the collection Ω of all the wind turbines ω to the set $[0, \infty)$ of all non-negative real numbers interpreted as 'times of failure.' Express the event A, that consists of all the wind turbines that have been functioning without a failure during the time period $[0, 10]$, in terms of T and ω's. (For a hint, see the definition of the subset L_x on page 48 of our textbook.)

Set 7 Suppose for the sake of this numerical example that the failure times of five windmills (for milling grains) are 6, 8, 9, 11, and 14 months. Let T be the random variable called 'failure time.'

(a) Calculate the probability $\mathbb{P}(7 \le T \le 11)$.

(b) Calculate the probability $\mathbb{P}(7 \le T < 11)$.

(c) Draw the distribution function $F_T(t) = \mathbb{P}(T \le t)$ for $t \ge 0$.

(d) Draw the reliability function $R_T(t) = \mathbb{P}(T > t)$ for $t \ge 0$.

Evaluation 1 (September 21, 2016)

(1) In a class there are 120 engineering students, 30 business students, and 50 science students. Assume that all students are viewed equally, and let none of them belong to more than one faculty. What is the probability of selecting an engineering or science student?

(2) Suppose that you have constructed a parallel system (hint: $W = W_1 \cup W_2 \cup W_3$) out of three independent components whose probabilities of functioning are $\mathbb{P}(W_1) = 0.9$, $\mathbb{P}(W_2) = 0.6$, and $\mathbb{P}(W_3) = 0.8$. What is the system's probability of functioning?

(3) Suppose that you have constructed a mixed system (hint: $W = W_1 \cap (W_2 \cup W_3)$) out of three independent components whose probabilities of functioning are $\mathbb{P}(W_1) = 0.9$, $\mathbb{P}(W_2) = 0.6$, and $\mathbb{P}(W_3) = 0.8$. What is the system's probability of functioning?

(4) Suppose that the failure times (outcomes of the failure random variable T) of five components are 14, 6, 8, 11, and 9. Draw the reliability function $R_T(t) = \mathbb{P}(T > t)$.

Set 8 (a) What is Bernoulli random variable?

(b) What is Bernoulli pmf (=probability mass function)?

(c) Can flipping a balanced coin be modeled using the Bernoulli random variable?

(d) Can flipping an unbalanced coin be modeled using the Bernoulli random variable?

Set 9 (a) What is binomial random variable?

(b) What is binomial pmf (=probability mass function)?

(c) Under what conditions can the binomial model be used?

(d) A certain type of windmill requires repairs once a month with probability 0.6. You own five windmills, and they are functioning independently of each other. What is the probability that you will not need to repair any of them during the upcoming month?

(e) A certain type of windmill requires repairs once a month with probability 0.6. You own five windmills, and they are functioning independently of each other. What is the probability that you will need to repair at least four of them during the upcoming month?

Set 10 (a) What is Poisson random variable?

(b) What is Poisson pmf (=probability mass function)?

(c) A certain type of windmill requires one repair a month in average. What is the probability that no repair will be required during the upcoming month?

(d) A certain type of windmill requires one repair a month in average. What is the probability that you will need to deal with at least two repairs during the upcoming month?

Set 11 (a) Let the random variable V count the number of heads when flipping two balanced coins. Calculate the probability mass function and draw the distribution function of V.

(b) Let the random variable W count the number of tails when flipping two balanced coins. Calculate the probability mass function and draw the distribution function of W.

(c) Let the random variable X count the number of working windmills in a farm of two identical and independent windmills, each having the reliability 0.9. Calculate the probability mass function and draw the distribution function of X.

(d) Let the random variable Y count the number of failed windmills in a farm of two identical and independent windmills, each having the reliability 0.9. Calculate the probability mass function and draw the distribution function of Y.

Set 12 (a) What is the uniform on $[0, 1]$ random variable?

(b) What is the uniform on $[0, 1]$ pdf (=probability density function)?

(c) What is the uniform on $[0, 1]$ distribution function?

(d) What is the uniform on $[0, 1]$ reliability function?

(e) If T is the uniform on $[0, 1]$ random variable, what is the probability that T is between $1/4$ and $3/4$?

Set 13 (a) What is the exponential random variable?

(b) What is the exponential pdf (=probability density function)?

(c) What is the exponential distribution function?

(d) What is the exponential reliability function?

(e) What is the exponential hazard rate function?

(f) If T is the exponential random variable with rate $\lambda = 2$, what is the probability that T is between 1 and 3?

Set 14 Let T be a random variable, whose distribution function we denote by F_T. With the notation $G_T(p) = \min\{t \ : \ F_T(t) \geq p\}$, the five-number summary is: min, 1st quartile $G_T(1/4)$, 2nd quartile/median $G_T(1/2)$, 3rd quartile $G_T(3/4)$, and max.

(a) Let T be the random variable taking values 0, 1, and 2 with probabilities $1/4$, $1/2$, and $1/4$, respectively. Calculate the five-number summary.

(b) Let T be the random variable taking each of the eight values 6, 8, 1, 3, 9, 5, 11, and 2 with the same probability $1/8$. Calculate the five-number summary.

(c) Let T be the uniform on $[0, 1]$ random variable. Calculate the five-number summary.

(d) Let T be the exponential random variable with the rate $\lambda = 1$. Calculate the five-number summary.

Set 15 (a) Suppose that the failure times of five windmills are 6, 8, 9, 11, and 14 months. Draw the reliability function, calculate the area under the reliability function, and compare it to the average value of the five failure times.

(b) Let X count the number of functioning windmills in a farm of two identical and independent windmills, each having the reliability 0.9. Calculate the expected/average/mean number of functioning windmills.

(c) If T is the uniform on $[a, b]$ random variable for some constants $a < b$, what is the expected/average/mean value μ_T of T?

(d) If T is the exponential random variable with the rate $\lambda = 2$, what is the expected/average/mean value $\mathbb{E}[T]$ of T?

(e) Prove that the expectation $\mathbb{E}[T]$ of a random variable $T \geq 0$ with density $f(t)$ can be written as the integral $\int_0^\infty R(t)dt$, usually called the mean time to failure (MTTF) in reliability engineering, where $R(t)$ is the reliability function.

(f) A certain type of windpump (for pumping groundwater into your house) fails according to the exponential distribution, whose reliability function is $R(t) = e^{-\lambda t}$, with time t measured in years. The windpump manual says that the MTTF is 5 years. What is the probability that the windpump works without a failure during the first two years after its installation?

Set 16 (a) Prove that the variance $\sigma^2 = \mathbf{E}[(X - \mu)^2]$ can be written as $\mathbf{E}[X^2] - \mu^2$ where $\mu = \mathbf{E}[X]$ is the mean of the random variable X.

(b) Let $g(c)$ be the function defined by the equation $g(c) = \mathbf{E}[(X - c)^2]$. Find the minimum of the function $g(c)$.

(c) Let X be the uniform on $[0, 1]$ random variable. Find its variance.

(d) Let X be the Bernoulli random variable. Find its standard deviation.

(e) Let X count the number of heads when flipping two balanced coins. Find the variance of X.

(f) Suppose that the failure times of three windmills are 2, 1, and 4 years. Find the standard deviation of the failure times.

We are now going into Chapter 4.

Set 17 In a class of 200 students, there are 150 students thinking of becoming engineers (event A), 100 students of becoming entrepreneurs (event B), and 50 students of becoming both engineers and entrepreneurs (event $A \cap B$).

(a) Given the data, do you conclude that thinking of becoming an engineer is independent of thinking of becoming an entrepreneur?

(b) What is the probability that the students who are thinking of becoming engineers are also thinking of becoming entrepreneurs?

(c) What is the probability that the students who are thinking of becoming entrepreneurs are also thinking of becoming engineers?

Set 18 Redundancy is frequently used to increase reliability. Suppose that two cables, 1 and 2, are used to transmit electricity from one point to another. Let A denote the event that cable 1 fails, and B that cable 2 fails. Based on past experience/data, we know that $\mathbf{P}(A) = 2/5$ and $\mathbf{P}(B) = 3/4$, and we also know that the probability of at least one cable failing is $\mathbf{P}(A \cup B) = 4/5$.

(a) Are the two cables functioning independently?

(b) What is the probability that both cables fail?

(c) What is the probability that cable 1 fails, given that cable 2 fails?

(d) What is the probability that cable 2 fails, given that cable 1 fails?

Set 19 A batch of 150 components were tested in a harsh environment. Some components passed the test but some failed due to high temperature or increased stress. The following failure-frequency table has been reported:

	Stress		
Temperature	PASS	FAIL	
PASS	40	20	60
FAIL	30	60	90
	70	80	150

(a) What is the probability of failure due to temperature?

(b) What is the probability of failure due to temperature given that the stress test has been passed?

(c) What is the probability of failure due to stress?

(d) What is the probability of failure due to stress given that the temperature test has been passed?

Evaluation 2 (October 5, 2016)

(1) Redundancy is frequently used to increase reliability. Suppose that two cables, 1 and 2, are used to transmit electricity from one point to another. Let A denote the event that cable 1 fails, and B that cable 2 fails. Based on past experience/data, we know that $\mathbf{P}(A) = 2/5$ and $\mathbf{P}(B) = 3/4$, and we also know that the probability of at least one cable failing is $\mathbf{P}(A \cup B) = 4/5$. What is the probability that cable 2 fails, given that cable 1 fails?

(2) Two events A and B are called independent if the probability $\mathbb{P}(A \cap B)$ is equal to the product of $\mathbb{P}(A)$ and $\mathbb{P}(B)$. Show that if the (non-empty) events A and B are disjoint, that is, if $A \cap B = \emptyset$, then the events must be dependent.

(3) What is random variable?

(4) A certain type of windmill requires repairs once a month with probability 0.6. You own five windmills, and they are functioning independently of each other. What is the probability that you will need to repair at least four of them during the upcoming month?

(5) Let T be a random variable, whose distribution function we denote by F_T. With the notation $G_T(p) = \min\{t \ : \ F_T(t) \geq p\}$, the five-number summary is: min, 1st quartile $G_T(1/4)$, 2nd quartile/median $G_T(1/2)$, 3rd quartile $G_T(3/4)$, and max. Let T be the random variable taking values 0, 1, and 2 with probabilities 1/4, 1/2, and 1/4, respectively. Calculate the five-number summary.

(6) Let X be the Bernoulli random variable. Derive its standard deviation.

(7) Let the random variable X count the number of working windmills in a farm of two identical and independent windmills, each having the reliability $p = 0.9$. Calculate the probability mass function and draw the reliability function of X.

(8) When does an event happen?

Set 20 Suppose[1] that $n = 5$ missiles have been fired towards a city, and each of them has probability p of evading each defensive layer of THAAD batteries: there might be one, two, or more layers. All shelters in the city can effectively withstand at most two hits.

(a) Suppose there is only one defensive layer of THAAD batteries, and let X be the number of incoming missiles that have evaded the defense. Assuming that $p = 0.05$, derive the probability mass function of X as well as the average number of missiles that would evade the defense. `(Hint: binomial distribution.)`

(b) Suppose there is only one defensive layer of THAAD batteries, and let X denote the number of incoming missiles that have evaded the defense. How small should

[1]This problem, which is of particular interest to civil and electronics engineers, has been inspired by the article entitled "THAAD: What It Can and Can't Do" by M. Elleman and M.J. Zagurek Jr. published in '*38 North*' on 10 March 2016. Web: `38north.org/2016/03/thaad031016/`

the parameter p be in order for the probability of more than two hits to be smaller than 0.001? `(Hint: binomial distribution.)`

(c) Suppose there are two defensive layers of THAAD batteries, and let X denote the number of incoming missiles that have evaded the two layers of defense. Assuming that $p = 0.05$, derive the probability mass function of X as well as the average number of missiles that would evade the defense. `(Hint: equation (C1) on p. 75 of our text and two applications of the binomial distribution.)`

(d) Suppose there are two defensive layers of THAAD batteries, and let X denote the number of incoming missiles that have evaded the two layers of defense. How small should the parameter p be in order for the probability of more than two hits to be smaller than 0.001? `(Hint: equation (C1) on p. 75 of our text and two applications of the binomial distribution.)`

Set 21 [2]The CIA World Factbook says that there are currently (approx.) 35 million Canadians and 324 million Americans. Furthermore, the leading market research company Ipsos MORI says that approximately 86% of Canadians and 83% of Americans are, generally speaking, happy. Suppose that you meet a happy North American, who could be either Canadian or American. What is the probability that the happy person is Canadian? `(Hint: you will need the Bayes Theorem on p. 75 of our text.)`

Set 22 A certain piece of equipment is regularly checked for failure at times $n = 1, 2, 3, \dots$, which could, for example, mean 'days.' Let $T \in \{1, 2, 3, \dots\}$ be the failure time, and let $h(n)$ denote the corresponding discrete hazard rate function.

(a) How is $h(n)$ defined in terms of the probability $\mathbb{P}$?

(b) Express $h(n)$ in terms of the reliability function of T.

(c) It is often more intuitive to first specify a hazard rate function and only then try to see how the corresponding reliability function looks like. For this reason, given a hazard rate function $h(n)$, derive a formula for the reliability function $R(n)$ in terms of $h(1), \dots, h(n)$.

(d) It is often more intuitive to first specify a hazard rate function and only then try to see how the corresponding probability mass function looks like. For this reason, given a hazard rate function $h(n)$, derive a formula for the probability mass function $\mathbb{P}(T = n)$ in terms of $h(1), \dots, h(n)$.

[2]You may wonder how this example (Set 21) might be related to engineering: Suppose that you are manufacturing certain systems and order certain components from two manufacturers, 1 and 2. From the second manufacturer, you usually ship, say, twice as many components than from the first one. Furthermore, the first manufacturer reports that about 86% of their components usually meet standards, whereas the second manufacturer reports 83%. (To make these low percentages more acceptable to you, you may think of grapes being sent to a winery; otherwise, change the percentages into higher ones and imagine, say, some mechanical or other components.) Suppose that just before installing a component into a system, you notice that the component is defective. You wonder if the defective component has more likely come from the first or the second manufacturer. For this, calculate and compare the two probabilities $\mathbb{P}(M_1 \mid D)$ and $\mathbb{P}(M_2 \mid D)$.

(e) Find $h(n)$ of the failure time T that follows the geometric distribution, whose probability mass function is $\mathbb{P}(T = n) = (1-c)^{n-1}c$ for all $n = 1, 2, 3, \ldots$, where $c \in (0, 1)$ is a constant. Draw the hazard rate function when $c = 1/2$.

Set 23 Let $T \geq 0$ be the 'failure time' random variable that follows the continuous distribution with a probability density function $f(t)$. Let $h(t)$ denote the hazard rate function of T.

(a) Express $h(t)$ in terms of the reliability function $R(t)$ only.

(b) It is often more intuitive to first specify a hazard rate function and only then try to see how the corresponding reliability function looks like. For this reason, given a hazard rate function $h(t)$, derive a formula for the reliability function $R(t)$. What is the reliability function of a hazard whose hazard rate function is constant, say $\lambda > 0$, that is, $h(t) = \lambda$ for all $t > 0$.

(c) It is often more intuitive to first specify a hazard rate function and only then try to see how the corresponding probability mass function looks like. For this reason, given a hazard rate function $h(t)$, derive a formula for the probability density function $f(t)$.

(d) Find $h(t)$ of the failure time T that follows the exponential distribution with rate $\lambda > 0$, that is, the distribution function of T is $F(t) = 1 - \exp\{-\lambda t\}$ for all $t > 0$. Draw the hazard rate function when $\lambda = 2$.

(e) Define the lack-of-memory property for any random variable T and prove that the property is satisfied by the exponential distribution with rate $\lambda > 0$.

(f) Find the hazard rate function $h(t)$ of the failure time T that follows the Weibull distribution with parameters $a > 0$ and $b > 0$, that is, the distribution function of T is $F(t) = 1 - \exp\{-at^b\}$ for all $t > 0$. Draw the hazard rate function when $a = 1$ and $b = 2$.

Set 24 (a) Let I_C denote the 'filter' (that is, the random variable) that takes the value 1 when event C occurs and 0 otherwise. Prove that if two events A and B are independent, then the expected value of the product $I_A I_B$ is the product of the expected values of I_A and I_B. (`Hint: page 96.`)

(b) Let X and Y be two discrete random variables. Prove that if these random variables are independent, then the expected value of the product XY is the product of the expected values of X and Y. (`Hint: page 101.`)

(c) Prove that if two random variables X and Y are independent, then the random variables are uncorrelated, that is, their covariance $\mathrm{Cov}(X, Y)$ is zero. (`Hint: page 102.`)

(d) Prove that if two random variables X and Y are independent, then the variance $\mathrm{Var}(X + Y)$ is the sum of the variances $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$. (`Hint: page 102.`)

Evaluation 3 (October 19, 2016)

(1) Let X be the Bernoulli random variable. Give the definition of this random variable and derive its variance.

(2) Let I_C denote the indicator of event C, that is, I_C takes value 1 when the event C occurs and 0 otherwise. Prove that if two events A and B are independent, then the expected value $\mathbb{E}[I_A I_B]$ of the product $I_A I_B$, defined by the equation $(I_A I_B)(\omega) = I_A(\omega) I_B(\omega)$ for all $\omega \in \Omega$, is the product of the expected values $\mathbb{E}[I_A]$ and $\mathbb{E}[I_B]$.

(3) Define the lack-of-memory property and show that it holds when the random variable T has the exponential reliability function $R(t) = e^{-\lambda t}$ with any rate $\lambda > 0$.

(4) Define the continuous-time hazard rate function $h(t)$ and then calculate it when the failure time T follows the Weibull distribution with parameters $a > 0$ and $b > 0$, that is, when the reliability function of T is $R(t) = \exp\{-at^b\}$ for all $t > 0$.

(5) Let T be the random variable that takes the values 0, 1, and 2 with the probabilities 1/4, 1/2, and 1/4, respectively. With F_T denoting its distribution function, the quantile function G_T of T is given by the equation $G_T(p) = \min\{t \ : \ F_T(t) \geq p\}$. Draw the distribution function F_T and calculate $G_T(1/3)$, $G_T(2/3)$, and $G_T(3/3)$.

(6) Two possibly load-sharing components make up a part of a system. It is known that the first component fails with probability 2/5, the second one fails with probability 3/4, and at least one of them fails with probability 4/5. What is the probability that the second component fails given that the first component has failed?

(7) Let $A \subset \Omega$ be an event. When does the event occur?

(8) Suppose that you usually order certain components from two manufacturers, and let your recent order contain three times more components from the second manufacturer than from the first one. The first manufacturer reports in their manual that about 86% of their components usually meet standards, whereas the second manufacturer reports 83%. Suppose that the shipments from the two manufacturers have already been unpacked and thus all the components are visually indistinguishable. Upon a close examination of a randomly selected component, you find it to be defective. What is the probability that the defective component comes from the first manufacturer?

Set 25 Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, whose common mean is μ and common variance σ^2. Let $\bar{X}$ denote the sample mean, that is, $\bar{X} = (X_1 + \cdots + X_n)/n$. (Hint: page 104.)

(a) Prove that $\mathbb{E}[\bar{X}] = \mu$.

(b) Prove that $\mathrm{Var}[\bar{X}] = \frac{\sigma^2}{n}$.

Set 26 Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, whose common mean is μ and common variance σ^2. With $\bar{X}$ denoting the sample mean, let

RSS be the residual sum of squares, that is, RSS $= \sum_{k=1}^{n}(X_k - \bar{X})^2$. `(Hint: page 105.)`

(a) Prove that RSS $= \left(\sum_{k=1}^{n} X_k^2\right) - n\bar{X}^2$.

(b) Prove that $\mathbb{E}[S^2] = \sigma^2$, where S^2 is the sample variance, which is defined by the equation $S^2 = \mathrm{RSS}/(n-1)$, that is, $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$.

Set 27 Let $T_1, \ldots, T_n$ be independent and identically distributed failure times, whose common reliability function is $R(t)$. Let $R_n(t)$ denote the empirical reliability function, which means that $R_n(t)$ is the proportion of those $T_1, \ldots, T_n$ that exceed t.

(a) Suppose that the observed failure times $T_1, \ldots, T_5$ of five identically manufactured wind turbines are 14, 6, 8, 11, and 9 months. Draw the empirical reliability function $R_n(t)$. `(Hint: the last problem of Evaluation 1.)`

(b) Let I_{B_k} denote the random variable that takes the value 1 when the event $B_k = \{\omega \in \Omega : T_k(\omega) > t\}$ occurs and 0 otherwise. Prove that I_{B_k} is a Bernoulli random variable, and express the probability of the event $I_{B_k} = 1$ in terms of the aforementioned common reliability function.

(c) Prove that $R_n(t)$ can be written as the sample mean $\frac{1}{n}\sum_{k=1}^{n} I_{B_k}$ of the Bernoulli random variables $I_{B_1}, \ldots, I_{B_n}$, where $B_k = \{\omega \in \Omega : T_k(\omega) > t\}$.

(d) Calculate the mean $\mathbb{E}[R_n(t)]$.

(e) Calculate the variance $\mathrm{Var}[R_n(t)]$.

Set 28 Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, whose common distribution function is $F(x)$. Let $F_n(x)$ denote the empirical distribution function, which means that $F_n(x)$ is the proportion of those $X_1, \ldots, X_n$ that do not exceed x. `(Hint: page 116.)`

(a) Let I_{A_k} denote the random variable that takes the value 1 when the event $A_k = \{\omega \in \Omega : X_k(\omega) \le x\}$ occurs and 0 otherwise. Prove that I_{A_k} is the Bernoulli random variable and express the probability of the event $I_{A_k} = 1$ in terms of the aforementioned common distribution function.

(b) Prove that $F_n(x)$ can be written as the sample mean $\frac{1}{n}\sum_{k=1}^{n} I_{A_k}$ of the Bernoulli random variables $I_{A_1}, \ldots, I_{A_n}$, where $A_k = \{\omega \in \Omega : X_k(\omega) \le x\}$.

(c) Calculate the mean $\mathbb{E}[F_n(x)]$.

(d) Calculate the variance $\mathrm{Var}[F_n(x)]$.

Set 29 (a) Prove Markov's Inequality, which says that, for every constant $\delta > 0$ and every non-negative random variable Y, the probability $\mathbb{P}(Y > \delta)$ does not exceed $\mathbb{E}(Y)/\delta$. `(Hint: page 106.)`

(b) Use Markov's Inequality to prove Tchebychev's Inequality, which says that, for every constant $\delta > 0$ and every random variable Z with mean μ and finite variance σ^2, the probability $\mathbb{P}(|Z - \mu| > \delta)$ does not exceed σ^2/δ^2. `(Hint: page 105.)`

(c) Weak Law of Large Numbers (WLLN): Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with common mean μ and (finite) common variance σ^2. Prove that the sample mean $\bar{X}$ approaches the (theoretical) mean μ in probability, meaning that no matter what constant $\delta > 0$ you choose, you have $\mathbb{P}(|\bar{X} - \mu| > \delta) \to 0$ when $n \to \infty$. `(Hint: page 107.)`

(d) Let $X_1, \ldots, X_n$ be independent Bernoulli random variables, each having the same probability of success p, that is, $p = \mathbb{P}(X_k = 1)$ for every k. Prove that the proportion of successes among these n Bernoulli random variables approaches the (theoretical) proportion p in probability when n gets larger and larger. `(Hint: page 106.)`

(e) Let $T_1, \ldots, T_n$ be independent and identically distributed failure times, whose common reliability function is $R(t)$. Let $R_n(t)$ denote the empirical reliability function. Prove that, for any fixed t, the random variable $R_n(t)$ approaches the proportion $R(t)$ in probability when n gets larger and larger.

(f) Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, whose common distribution function is $F(x)$. Let $F_n(x)$ denote the empirical distribution function. Prove that, for any fixed t, the random variable $F_n(t)$ approaches the proportion $F(t)$ in probability when n gets larger and larger.

This completes our study of Chapter 4.

Evaluation 4 (November 2, 2016)

(1) Let $T \geq 0$ be a failure-time random variable. Prove that the probability $\mathbb{P}(T > 2)$ is equal to the expectation of the indicator random variable $1_{\{T>2\}}$.

(2) Prove Markov's Inequality, which says that, for every constant $\delta > 0$ and every non-negative random variable Y, the probability $\mathbb{P}(Y > \delta)$ does not exceed $\mathbb{E}(Y)/\delta$.

(3) A batch of 150 components were tested in a harsh environment. Some components passed the test but some failed due to high temperature or increased stress. The following failure-frequency table has been reported:

	Stress		
Temperature	PASS	FAIL	
PASS	40	20	60
FAIL	30	60	90
	70	80	150

Calculate the probability of failure due to temperature or stress.

(4) Give the formula of the uniform on $[0, 1]$ probability density function (p.d.f.) and then derive its distribution function (d.f.).

(5) Let the random variable X count the number of heads when flipping two balanced coins. Calculate the probability mass function and draw the distribution function of this random variable.

(6) Let T be a failure-time random variable, and let $g(x)$ be the function defined by the equation $g(x) = \mathbf{E}[(T - x)^2]$. Find the minimum of this function.

(7) Let $T \geq 0$ be the failure-time random variable that follows the continuous distribution with a probability density function $f(t)$. Suppose that someone has given you an expression of the reliability function $R(t)$. How would you then calculate the hazard rate function $h(t)$ of T?

(8) The CIA World Factbook says that there are currently (approx.) 35 million Canadians and 324 million Americans. Furthermore, the leading market research company Ipsos MORI says that approximately 86% of Canadians and 83% of Americans are, generally speaking, happy. Suppose that you meet a happy North American, who could be either Canadian or American. What is the probability that the happy person is Canadian?

We are now diving simultaneously into Chapters 5 and 6.

Set 30 Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with the same mean μ and variance σ^2, and let $\bar{X}$ denote the sample mean. Tchebychev's inequality says that the probability $\mathbb{P}(|\bar{X} - \mu| > \delta)$ does not exceed $\mathrm{Var}(\bar{X})/\delta^2$ for every constant $\delta > 0$, where $\mathrm{Var}(\bar{X})$ denotes the variance of $\bar{X}$.

(a) Prove that the statement

$$|\bar{X} - \mu| > \sqrt{20}\frac{\sigma}{\sqrt{n}}$$

holds with a probability that does not exceed 0.05.

(b) Prove that the statement

$$\bar{X} - \sqrt{20}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \sqrt{20}\frac{\sigma}{\sqrt{n}} \tag{1}$$

holds with a probability that is not smaller than 0.95. NOTE: Statement (1) gives the following confidence interval

$$\left[\bar{X} - \sqrt{20}\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X} + \sqrt{20}\frac{\sigma}{\sqrt{n}}\right] \tag{2}$$

for the (unknown) population mean μ, with a confidence level at least 95% and for every sample size n.

(c) Prove that the statement

$$\mu - \sqrt{20}\frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu + \sqrt{20}\frac{\sigma}{\sqrt{n}}$$

holds with a probability that is not smaller than 0.95.

(d) Let $\omega_1^{\mathrm{act}}, \ldots, \omega_n^{\mathrm{act}}$ be a sample from a population Ω, and let $X : \Omega \to \mathbf{R}$ be a filter (that is, a random variable in Statistics, or a measurable function in Mathematics) that produces n outputs (called observations in Statistics) $x_1^{\mathrm{obs}}, \ldots, x_n^{\mathrm{obs}}$ by the formula $x_i^{\mathrm{obs}} = X(\omega_i^{\mathrm{act}})$. Let $\bar{x}$ denote the average of $x_1^{\mathrm{obs}}, \ldots, x_n^{\mathrm{obs}}$. Does the confidence interval

$$\left[\bar{x} - \sqrt{20}\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + \sqrt{20}\frac{\sigma}{\sqrt{n}}\right] \tag{3}$$

cover the (unknown) population mean μ or not?

Set 31 Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with the same mean μ and variance σ^2, and let $\bar{X}$ denote the sample mean. Use the Central Limit Theorem (CLT, page 156 in our textbook) to establish the following (asymptotic when $n \to \infty$) statements:

(a) The statement

$$|\bar{X} - \mu| > 1.96\frac{\sigma}{\sqrt{n}}$$

holds with the probability 0.05.

(b) The statement

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.96\frac{\sigma}{\sqrt{n}} \tag{4}$$

holds with the probability 0.95. NOTE: Statement (1) gives the following (asymptotic when $n \to \infty$) 95% confidence interval

$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right] \tag{5}$$

for the (unknown) population mean μ.

(c) The statement

$$\mu - 1.96\frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu + 1.96\frac{\sigma}{\sqrt{n}}$$

holds with the probability 0.95.

(d) Let $\omega_1^{\text{act}}, \dots, \omega_n^{\text{act}}$ be a sample from a population Ω, and let $X : \Omega \to \mathbf{R}$ be a filter (that is, a random variable in Statistics, or a measurable function in Mathematics) that produces n outputs (called observations in Statistics) $x_1^{\text{obs}}, \dots, x_n^{\text{obs}}$ by the formula $x_i^{\text{obs}} = X(\omega_i^{\text{act}})$. Let $\bar{x}$ denote the average of $x_1^{\text{obs}}, \dots, x_n^{\text{obs}}$. Does the 95% confidence interval

$$\left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right] \tag{6}$$

cover the (unknown) population mean μ or not?

Set 32 Suppose that you own a wind farm with five wind turbines of the same type, and suppose that you have kept the records of their first failure times $t_1^{\text{obs}}, \dots, t_5^{\text{obs}}$, which are 14, 6, 8, 11, and 9 months.

(a) Draw the sample (empirical) reliability function.

(b) Calculate the area under the empirical reliability function.

(c) Calculate the sample mean $\bar{t}$ of the data.

(d) Construct a 95% (asymptotic) confidence interval for the population MTTF (mean time to failure) when someone has told you that, based on historical records, the population variance σ^2 is 9.

(e) Calculate the sample standard deviation $\widehat{\sigma}$ of the data.

(f) Construct a 95% (asymptotic) confidence interval for the population MTTF when no information about the population variance σ^2 is available.

Set 33 Suppose that you have tossed a coin – which may or may not be balanced – five times and reported the outcomes 1, 0, 0, 1, and 1, with Heads coded by 1's and Tails by 0's.

(a) Calculate the sample proportion $\widehat{p}$ of Heads.

(b) Construct a 95% (asymptotic) confidence interval for the population probability of Heads.

(c) Construct a *conservative* 95% (asymptotic) confidence interval for the population probability of Heads.

Set 34 Suppose that you own a wind farm with five wind turbines of the same type, and suppose that you have kept the records of their first failure times $t_1^{\text{obs}}, \dots, t_5^{\text{obs}}$, which are 14, 6, 8, 11, and 9 months.

(a) Construct a 95% (asymptotic) confidence interval for the probability that the wind turbine of this type functions without a failure for $t = 10$ months after its installation.

(b) Construct a *conservative* 95% (asymptotic) confidence interval for the probability that the wind turbine of this type functions without a failure for $t = 10$ months after its installation.

(c) Construct a 95% (asymptotic) confidence interval for the probability that the wind turbine of this type fails within $t = 10$ months of its installation.

(d) Construct a *conservative* 95% (asymptotic) confidence interval for the probability that the wind turbine of this type fails within $t = 10$ months of its installation.

Set 35 Suppose that you have designed a product and want to assess its reliability in the form of its functionality during the required, say, 10 day mission. For this, you want to have an estimate of the reliability function $R(t)$ at $t = 10$, and you also want to have a 95% (asymptotic) confidence interval for $R(10)$ whose margin of error does not exceed 0.01. How many prototypes should you manufacture and then wait for their failure times in order to construct the aforementioned confidence interval when:

(a) there is no additional information available to you?

(b) your expertise in physics, chemistry, and of course engineering tells you that $R(10)$ should be somewhere between 0.85 and 0.98?

(c) your expertise tells you that $R(10)$ should be somewhere between 0.45 and 0.70?

Set 36 Suppose that you own a fleet of taxis consisting of hybrid cars of the same type, and suppose that you have decided to re-invest some of your profits by adding solar roofs to the taxis. Hence, you buy the best solar cells available on the market and, being an innovative engineer, decide to arrange them in a most efficient way. You quickly realize that the shape of the car roof is one of the most important factors, and you can easily shape the roof into any form of your liking (the drag coefficient is not an issue for your taxis because they cannot drive fast in the city). Your computer modeling suggests a certain shape, which is somewhat different from the original roof-shape of the taxis. Should you keep the original roof-shape or modify it? To answer this question, you randomly select ten taxis from your fleet, add solar roofs to all of them, drive for a week, and record their gas-free mileage. After that, you modify the roofs of the same ten taxis, add solar roofs, drive them for a week, and record their gas-free mileage. Hence, you have ten pairs of data. Construct a 95% (asymptotic) confidence interval for the difference between the (population) average gas-free mileage under the original and modified roofs.

Set 37 Suppose that instead of using ten same cars as in Set 36, you are now dealing with a taxi fleet that is sufficiently large to allow you to select a number of cars, say 10, whose unmodified roofs are fitted with solar panels, and also select a number of cars, say 15, whose modified roofs are fitted with solar panels. Hence, the whole experiment can now be run within one week, instead of two as in Set 36, and so you obtain two sets of data at the same time: one has 10 and another 15 observations (gas-free mileage recordings). Construct a 95% (asymptotic) confidence interval for the difference between the (population) average gas-free mileage under the original and modified roofs.

Set 38 Let $x_1^{\text{obs}}, \ldots, x_n^{\text{obs}}$ be observations, and let $\overline{x}$ and $\widehat{\sigma}^2$ be the sample mean and variance, respectively. Does the 95% confidence interval

$$\left[\overline{x} - 1.96\frac{\widehat{\sigma}}{\sqrt{n}}\ ,\ \overline{x} + 1.96\frac{\widehat{\sigma}}{\sqrt{n}}\right] \tag{7}$$

(a) cover the (unknown) population mean μ?

(b) cover the sample mean $\overline{x}$?

Set 39 Suppose you have paired data $(x_1^{\text{obs}}, y_1^{\text{obs}}), \ldots, (x_n^{\text{obs}}, y_n^{\text{obs}})$, which are outcomes of two random variables X and Y. You construct a 95% confidence interval for the difference of the means of X and Y.

(a) Assuming that the correlation between X and Y is positive, is your constructed confidence interval shorter or longer than the 95% confidence interval that you would have in the case of independent X and Y? Prove your answer.

(b) If the 95% confidence interval does not cover 0, would you retain or reject someone's claim that the means of X and Y are equal? Justify your answer.

(c) If the entire 95% confidence interval is to the right of 0, would you conclude or not that the mean of X is larger than the mean of Y? Justify your answer.

Set 40 Least-squares regression line

(a) Derive expressions for a and b that give the minimal value of the two-argument function defined by $g(a, b) = \mathbb{E}[(a + bX - Y)^2]$. (Hint: a is called the intercept and b the slope of the least-squares regression line, which is $y = a + bx$.)

(b) Construct empirical estimators $\widehat{a}$ and $\widehat{b}$ for a and b, respectively, and then compute the estimators and draw the corresponding least-squares regression line using the following data (Aven, 2010, p. 47):[3]

Year (i)	1	2	3	4	5	6	7	8	9	10	1	12
Market (x_i)	0.15	0.13	0.07	0.12	-0.04	0.31	0.23	0.31	0.02	-0.07	0.07	0.02
Fund (y_i)	-0.05	0.05	0.01	0.25	0.04	0.15	0.40	0.29	0.33	-0.03	0.02	-0.02

(c) Prove the equation

$$\frac{\mathrm{cov}(X, Y)}{\sigma_X^2} = \mathrm{corr}(X, Y)\frac{\sigma_Y}{\sigma_X}, \tag{8}$$

[3] Aven, T. (2010). Misconceptions of Risk. Wiley, Chichester, UK.

where $\mathrm{cov}(X, Y)$ is the covariance between X and Y, and $\mathrm{corr}(X, Y)$ is the correlation between X and Y. (Note: The left-hand side of equation (8) is the famous "beta" that every financial portfolio manager knows and uses on the daily basis, with X being the market return and Y the fund return.)

(d) Prove that the mean squared error $\mathrm{MSE} := \mathbb{E}[(\widehat{Y} - Y)^2]$ between the least squares predictor $\widehat{Y} = a + bX$ of the response variable Y is equal to $\sigma_Y^2(1 - \rho^2)$, where σ_Y^2 is the variance of Y and $\rho = \mathrm{corr}(X, Y)$ is the correlation between X and Y.

Set 41 Let X have the uniform on $[-1, 1]$ density, and let Y be another random variable given by the equation $Y = X^2$. Hence, the value of Y is completely determined by the value of X. Are the two random variables X and Y correlated or uncorrelated? Prove your answer.

Set 42 Suppose that you look after a wind farm with 20 wind turbines, among which 5 turbines are in location A and the remaining 15 turbines are in another location B. Calculate the mean time to failure of the entire wind farm when you only know that the mean time to failure of the turbines in location A is 6 months, and in location B is 4 months.

Set 43 Suppose that you own a grocery store. Let N denote the number of shoppers who visit your grocery store, and let X_n (where $n = 1, 2, \dots$) denote the amount of money that the n-th shopper spends during one week. Based on historical data, you know that each shopper spends, in average, \$ 200 in your store during one week, and the average number of shoppers who visit your store during one week is 500. Let S_N be the total sum of money spent by all the N shoppers during one week. Assume that the number of shoppers N and the amounts X_n, $n \geq 1$, of money that they spend are independent. What is the average $\mathbb{E}[S_N]$ of the total amount of money S_N spent during one week at your grocery store?

Set 44 Let X be a discrete random variable, and let A and B be two disjoint (that is, $A \cap B = \emptyset$) and exhaustive (that is, $A \cup B = \Omega$) subsets of the sample space Ω. Prove

$$\mathbb{E}[X] = \mathbb{E}[X \mid A]\mathbb{P}(A) + \mathbb{E}[X \mid B]\mathbb{P}(B). \tag{9}$$

Set 45 Let X and Y be two discrete random variables. Prove that the expectation of the conditional expectation is the unconditional expectation, that is,

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]. \tag{10}$$

Set 46 The mean residual life function (MRL) defined by $\mathrm{MRL}(t) = \mathbb{E}[T - t \mid T > t]$ is as important in reliability engineering as the hazard rate function.

(a) Construct an empirical estimator for the function $\mathrm{MRL}(t)$ and then calculate its values at $t = 7$ and $t = 10$ when the observed failure times $t_1^{\mathrm{obs}}, \dots, t_5^{\mathrm{obs}}$ are 14, 6, 8, 11, and 9 months.

(b) Let $R(t)$ denote the reliability function of the failure-time random variable T. Prove the equation

$$\mathrm{MRL}(t) = \frac{1}{R(t)} \int_t^\infty R(x)dx. \tag{11}$$

(c) Assume that the failure-time random variable T has a probability density function $f(t)$, and let $R(t)$ denote the reliability function of T. Prove the equation

$$\mathrm{MRL}(t) = \frac{1}{R(t)} \int_t^\infty (x-t)f(x)dx. \tag{12}$$

(d) Calculate the mean residual life function $\mathrm{MRL}(t)$ of the exponential random variable T with the rate parameter $\lambda > 0$, that is, its reliability function is $R(t) = e^{-\lambda t}$ for all $t \geq 0$.