STK-MAT3710/4710: Probability Theory
Mandatory assignment Fall 2024

Submission deadline

Thursday 17th October 2024, 14:30 in Canvas (`canvas.uio.no`).

Instructions

If you are taking STK-MAT3710, you can choose between scanning handwritten notes or typing the solution directly on a computer (for instance with LaTeX). If you are taking STK-MAT4710, you must write your solution in LaTeX. The assignment must be submitted as a single pdf file. Scanned pages must be clearly legible. The submission must contain your name and course code.

It is expected that you give a clear presentation with all necessary explanations. Remember to include all relevant plots and figures. You only have **one attempt** at the assignment, and you need to have the assignment approved in order to take the exam. All aids, including collaboration, are allowed, but the submission must be written by you and reflect your understanding of the subject. If we doubt that you have understood the content you have handed in, we may request that you give an oral account.

Answer all problems or, at least, try. To pass the assignment we expect that you answer 2 out of the 3 exercises correctly.

Complete guidelines about delivery of mandatory assignments:

`uio.no/english/studies/admin/compulsory-activities/mn-math-mandatory.html`

The solution of the assignment is given in blue.

Problem 1

Write down and explain with **at least one** example the concepts presented below. It is okay to use examples from the lecture notes, but it will be **highly** appreciated if you build your own examples!

(a) A σ-algebra on a set E.

Definition 0.1 (σ algebra) *Let $\mathcal{E} \subseteq \mathcal{P}(E)$ be a non-empty family of subsets of E. Then we say that $\mathcal{E}$ is a* σ-algebra *of sets if it contains the total set E, is closed under countable unions and complements. In other words,*

1. *$E \in \mathcal{E}$,*
2. *given $\{E_i\}_{i=1}^{\infty} \subset \mathcal{E}$ implies that $\bigcup_{i=1}^{\infty} \in \mathcal{E}$,*
3. *given $F \in \mathcal{E}$ implies that $F^c = E \setminus F \in \mathcal{E}$.*

Example 0.2 *Let $E = \{a, b, c\}$ then three possible σ-algebras on E are*

$$\mathcal{E}_1 = \{\emptyset, \{a\}, \{b, c\}, E\},$$
$$\mathcal{E}_2 = \{\emptyset, \{c\}, \{a, b\}, E\},$$
$$\mathcal{E}_3 = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, E\}.$$

We see that $\mathcal{E}_1 \subset \mathcal{E}_3$ and $\mathcal{E}_2 \subset \mathcal{E}_3$ but $\mathcal{E}_1$ and $\mathcal{E}_2$ are not contained in each other.

(b) A measurable space.

Definition 0.3 (Measurable space) *The pair $(E, \mathcal{E})$ where E is a set and $\mathcal{E}$ is a σ-algebra of subsets of E is called a* measurable space.

Example 0.4 *Any of the above examples are measurable spaces $(E, \mathcal{E}_i)$ for $i = 1, 2, 3$.*

(c) A measure μ on a measurable space and a measure space.

Definition 0.5 (Measure) *Let $(E, \mathcal{E})$ be a measurable space. Then a mapping $\mu : \mathcal{E} \to [0, \infty]$ is called a* measure *if:*

1. $\mu(\emptyset) = 0$,
2. *(σ-additivity): given $\{E_i\}_{i=1}^{\infty} \subset \mathcal{E}$ such that $E_i \cap E_j = \emptyset$, $i \neq j$, then*

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i).$$

Definition 0.6 (Measure space) *A* measure space *is a triplet $(E, \mathcal{E}, \mu)$ formed by a set E, a σ-algebra $\mathcal{E}$ on E and a measure μ on the measurable space $(E, \mathcal{E})$.*

Example 0.7 *Some examples are:*

1. **Borel measures:** *Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be the measurable space consisting of the real numbers and the Borel σ-algebra. Then any measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called* Borel measure. *Example:* **Lebesgue measure** *is a Borel measure where $\mu((a, b]) = b - a$.* **Lebesgue-Stieltjes measure** *is a Borel measure where $\mu_F(a, b]) = F(b) - F(a)$ for a fixed increasing function F.* **Dirac measure** *is a Borel measure such that $\mu_a(A) = 1$ if $a \in A$ and $\mu_a(A) = 0$ if $a \notin A$ for a fixed point $a \in \mathbb{R}$ and all Borel sets A.*
2. **Counting measure:** *The measure that counts the number of elements of any subset is known as the* counting measure, *i.e. $\mu(F) = \#F$ for all $F \in \mathcal{E}$. For example, if $E = \mathbb{N}$ and $\mathcal{E} = \mathcal{P}(\mathbb{N})$ then $\mu(\{1, \ldots, n\}) = n$ for any $n \in \mathbb{N}$ and $\mu(\mathbb{N}) = \infty$. This measure is σ-finite, but not finite.*
3. **Probability measure:** *Any measure μ in a measure space $(E, \mathcal{E}, \mu)$ such that $\mu(E) = 1$ is called a* probability measure *and $(E, \mathcal{E}, \mu)$ is called a* probability space.

(d) A complete measure (with an example of complete and non-complete measure).

Definition 0.8 (Negligible set) *Let $(E, \mathcal{E}, \mu)$ be a measure space and $N \subseteq E$. We say that N is a* negligible set *if there exists $F \in \mathcal{E}$ such that $N \subseteq F$ and $\mu(F) = 0$. We denote by $\mathcal{N}_\mu$ or simply $\mathcal{N}$ when it clear what the measure is, the set of all negligible sets.*

Definition 0.9 (Complete measure) *Let $(E, \mathcal{E}, \mu)$ be a measure space and denote by $\mathcal{N}_\mu$ the set of negligible sets. We say that μ is* complete *when all the negligible sets are measurable, that is $\mathcal{N}_\mu \subseteq \mathcal{E}$. In other words, for every $F \subseteq N$, $N \in \mathcal{N}_\mu$ implies that $F \in \mathcal{N}_\mu$.*

Example 0.10 *An example of an incomplete measure is the following: $E = \{a, b, c\}$ with the σ-algebra $\mathcal{E} = \{\emptyset, \{a, b\}, \{c\}, E\}$ and a measure such that $\mu(\{a, b\}) = 0$. Then μ is not complete because $\{a\}$ is a negligible set, indeed, $\{a\} \subset \{a, b\} \in \mathcal{E}$ with $\mu(\{a, b\}) = 0$ but $\{a\}$ itself is not measurable since $\{a\} \notin \mathcal{E}$. Nevertheless, we could «complete» the measure μ by adding $\{a\}$ into $\mathcal{E}$, this would force us to also add the set $\{b, c\}$ and hence also $\{b\} = \{a, b\} \cap \{b, c\}$ and thus $\{a, c\}$ as well. Now, $\mu(\{a\}) \leq \mu(\{a, b\})$ since $\{a\} \subset \{a, b\}$ and hence $\mu(\{a\}) = 0$ since $\mu(\{a, b\}) = 0$. This would also imply $\mu(\{b, c\}) = \mu(E)$ since $\mu(\{b, c\}) = \mu(\{a, b, c\} \setminus \{a\}) = \mu(E) - \mu(\{a\}) = \mu(E)$.*

(e) A measurable function between measurable spaces.

Definition 0.11 (Measurable mapping) *Let $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ be a mapping between measurable spaces. We say that f is $(\mathcal{E}, \mathcal{F})$-*measurable *if for every $B \in \mathcal{F}$ we have that $f^{-1}(B) \in \mathcal{E}$. If the σ-algebras are clear from the context we simply say that f is* measurable.

Example 0.12 *Let $(E, \mathcal{E})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable spaces. Let $A \subseteq E$ be a subset of E. Then consider the function $\mathbb{I}_A : E \to \mathbb{R}$ that assigns 1 to elements inside A and 0 to elements outside A. Then $\mathbb{I}_A$ is measurable if, and only if $A \in \mathcal{E}$. Indeed, let $B \in \mathcal{B}(\mathbb{R})$ be a Borel set. Then if $0 \notin B$ and $1 \notin B$, we have $\mathbb{I}_A^{-1}(B) = \emptyset \in \mathcal{E}$. If B is such that $1 \in B$ but $0 \notin B$ then $\mathbb{I}_A^{-1}(B) = A$. If B is such that $1 \notin B$ and $0 \in B$ the $\mathbb{I}_A^{-1}(B) = A^c$ and finally, if $1 \in B$ and $0 \in B$ then $\mathbb{I}_A^{-1}(B) = E \in \mathcal{E}$. Hence, $\mathbb{I}_A$ is measurable if, and only if $A \in \mathcal{E}$ (and hence $A^c \in \mathcal{E}$ by definition of σ-algebra).*

(f) The concept of almost everywhere (a.s) or almost surely (a.s) in the context of probability. Give an example of a property that holds a.e. or a.s. but not on the entire space.

Definition 0.13 (Almost everywhere) *Let $(E, \mathcal{E}, \mu)$ be a measure space. A property P is said to hold* almost everywhere *in E if there exists a set $N \in \mathcal{E}$ with $\mu(N) = 0$, and all $x \in E \setminus N$ have the property P. Another common way of expressing the same thing is to say that «almost every point satisfies P», or that «for almost every x, $P(x)$ holds».*

Definition 0.14 (Almost surely) *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, that is a measure space with $\mathbb{P}[\Omega] = 1$. Then we say that a property P holds* almost surely *if it holds in all points ω except on a set $N \in \mathcal{A}$ such that $\mathbb{P}[N] = 0$.*

Example 0.15 *Let $f : \mathbb{R} \to \mathbb{R}$ be the function $f(x) = |x|$. Then f is C^∞ (Lebesgue) a.e. Indeed, the only point where f is not C^∞ is at $x = 0$ but $\lambda(\{x\}) = 0$ (where λ here denotes the Lebesgue measure), hence f is C^∞* almost everywhere. *The same would be true if the property failed in a countable set of points.*

(g) The concept of push-forward measure.

Definition 0.16 (Push-forward measure) *Let $(E, \mathcal{E}, \mu)$ be a measure space, $(F, \mathcal{F})$ a measurable space and $f : E \to F$ a measurable transformation. Then the transformation defined by*

$$\mu_f(A) \triangleq \mu \circ f^{-1}(A) = \mu(f^{-1}(A)), \quad A \in \mathcal{F},$$

defines a measure on $(F, \mathcal{F})$. The measure μ_f is called the push-forward measure *via f. Sometimes it is denoted by $f\#\mu$.*

Example 0.17 *Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ a measure space and $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$. Then $\mu_f = f\#\mu = \mu \circ f^{-1}$ (the first two notations are just notations for the same thing) is the measure acting on Borel sets A as $\mu_f(A) = 2\lambda(A)$.*

From now on we consider real-valued functions on general measurable spaces, i.e. $f : E \to \mathbb{R}$. Define and explain with at least **one** example the following concepts:

(h) A simple function.

Definition 0.18 (Simple function) *Let $(E, \mathcal{E}, \mu)$ be a measure space and $\{E_i\}_{i=1}^n \subset \mathcal{E}$ a partition of E, i.e. $\cup_{i=1}^n E_i = E$, $E_i \in \mathcal{E}$, with $E_i \cap E_j = \emptyset$, $i \neq j$. Moreover, let $\{a_i\}_{i=1}^n \subset \mathbb{R}$, $a_i > 0$, a sequence of positive real numbers. We say that a function $f : E \to \mathbb{R}$ is* simple *if it is of the form:*

$$f(x) = \sum_{i=1}^{n} a_i \mathbb{I}_{E_i}(x), \quad x \in E.$$

Example 0.19 *On $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, λ Lebesgue measure, then any constant function is a simple function or*

$$f(x) = \sum_{i=1}^{n} a_i \mathbb{I}_{(i-1,i]}(x),$$

for a sequence of non-negative numbers a_i.

(i) The integral of a simple function with respect to a measure.

Definition 0.20 (Integral w.r.t. μ for simple functions) *Let f be a simple function on a measure space $(E, \mathcal{E}, \mu)$ (as in Definition 0.18). Then the integral of f with respect to the measure μ is defined as*

$$\int_E f d\mu \triangleq \sum_{i=1}^{n} a_i \mu(E_i).$$

Example 0.21 *On $(E, \mathcal{E}, \mu)$ a measure space and $f : E \to \mathbb{R}$ the constant function $f(x) = 1$ for all x then we have for all measurable sets $A \in \mathcal{E}$:*

$$\int_A f(x) d\mu(x) = \int_A d\mu(x) = \mu(A).$$

Or if given a measurable set A, we have the function

$$f(x) = a_1 \mathbb{I}_A(x) + a_2 \mathbb{I}_{A^c}(x),$$

for non-negative numbers a_1 and a_2 then

$$\int_E f(x) d\mu(x) = a_1 \mu(A) + a_2 \mu(A^c).$$

(j) The integral of a measurable function with respect to a measure.

Definition 0.22 (Integral of non-negative measurable functions) *Let $f : E \to \mathbb{R}$ be a non-negative measurable function. Let $\{f_n\}_{n\geq 1}$ be a sequence of simple functions such that $\lim_n f_n = f$ pointwise. Then, we define the integral of f with respect to the measure μ as*

$$\int_E f d\mu \triangleq \lim_{n\to\infty} \int_E f_n d\mu.$$

Then by splitting any function f into its positive and negative parts, $f = f^+ - f^-$ where f^+ and f^- are non-negative, we can apply the definition above to f^+ and f^- in order to define the integral of f.

Definition 0.23 (Integral of a measurable function w.r.t. μ) *Let $f : E \to \mathbb{R}$ be a measurable function and let f^+ and f^- denote its positive and negative part, respectively. Then if $\int_E f^+ d\mu < \infty$ and $\int_E f^- d\mu < \infty$, we say that f is μ-integrable and*

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu.$$

In some cases the Lebesgue integral can be understood and computed as a Riemann integral. So

$$\int_{[0,1]} x d\lambda(x) = \int_0^1 x dx,$$

where the first one is "understood" as a Lebesgue integral and defined as in Definition 0.23 and the second one is approached in a completely different way: taking a partition of $[0, 1]$, writing Riemann sums and taking the limit of the Riemann sums. So, the definition of Lebesgue integral is rather theoretical and little applicable in some cases. But whenever we can connect it to the Riemann integral, then we can say:

$$\int_{[0,1]} x d\lambda(x) = \int_0^1 x dx = \frac{x^2}{2}\Big|_{x=0}^{x=1} = \frac{1}{2},$$

where the "computing rule" we used is the typical for Riemann integrals, also known as Newton–Leibniz theorem[1].

(k) The integral of a function with respect to the push-forward measure.

[1] Newton–Leibniz theorem is just the name of the rule you use for computing elementary integrals, you did maybe not know that is had a name! Some people also call it Barrow's rule in honour of Isaac Barrow, teacher of Isaac Newton.

Theorem 0.24 (Integration w.r.t. pushforward measure) *Let $(E, \mathcal{E}, \mu)$ be a measure space, $(F, \mathcal{F})$ a measurable space and $f : E \to F$ a measurable transformation. Consider a real-valued function $g : F \to \mathbb{R}$. Then the integral of g on $(F, \mathcal{F}, f\#\mu)$ exists if, and only if the integral of $g \circ f$ exists on $(E, \mathcal{E}, \mu)$ and it holds that*

$$\int_F g d(\mu \circ f^{-1}) = \int_E g \circ f d\mu.$$

Example 0.25 *The most important example of this result for our purposes is the connection between integral w.r.t. a probability measure and the integral on the real numbers. This is,*

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x d\mu_X(x),$$

where μ_X is the law of X. Recall that the law of a r.v. X is defined as $\mu_X(B) = \mathbb{P}[X \in B]$, for Borel sets B. Then, we can connect the measure theoretical concept of integral with respect to the measure μ_X with the integral w.r.t. the function such that $\mu_X((a, b]) = F_X(b) - F_X(a)$ and F_X is what we call distribution function *of X and $F_X(x) = \mathbb{P}[X \leq x]$, $x \in \mathbb{R}$. Thus,*

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x d\mu_X(x) = \int_{\mathbb{R}} x dF_X(x),$$

where the first integral is an integral w.r.t. the measure $\mathbb{P}$, while the second is an integral w.r.t. the measure μ_X and the last integral is an integral w.r.t. the function *F_X in the Riemann-Stieltjes sense.*

Now that we have covered the most important concepts on measure theory, we focus on probability theory. You can use the concepts defined above in order to define the following ones:

(l) Concept of probability, probability space and random variable.

A probability measure is simply a measure such that $\mu(E) = 1$ but we give the definition again:

Definition 0.26 (Probability measure) *Let Ω be a sample space and $\mathcal{A}$ a σ-algebra of events of Ω. Then a* probability measure *$\mathbb{P}$ is a mapping $\mathbb{P} : \mathcal{A} \to [0, 1]$ satisfying the following conditions:*

1. $\mathbb{P}[\Omega] = 1$.

2. *(σ-additivity) if $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$ is a countable sequence of pairwise disjoint events, i.e. $A_i \cap A_j = \emptyset$ for all $i \neq j$, then*

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} \mathbb{P}[A_n].$$

A probability space is simply a measure space $(E, \mathcal{E}, \mu)$ where the measure μ is a probability measure. A random variable is simply a measurable mapping from Ω to $\mathbb{R}$.

Definition 0.27 (Random variable) *A* random variable X *is a mapping between Ω and $\mathbb{R}$, i.e. $X : \Omega \to \mathbb{R}$ such that for every $x \in \mathbb{R}$,*

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}.$$

The general definition should be that measurable sets from $\mathbb{R}$, that is Borel sets B are mapped to events in $\mathcal{A}$, i.e. $X^{-1}(B) \in \mathcal{A}$. But since the sets of the form $(-\infty, x]$, $x \in \mathbb{R}$ are Borel sets that generate the whole σ-algebra $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ then it is sufficient to check that $X^{-1}((-\infty, x]) \in \mathcal{A}$ for all x, hence the definition given in this form. The set $X^{-1}((-\infty, x])$ is exactly $\{\omega \in \Omega : X(\omega) \leq x\}$.

Example 0.28 *Any measure space $(E, \mathcal{E}, \mu)$ with $\mu(E) = 1$ is a probability. So, $([0,1], \mathcal{B}([0,1]), \lambda)$ is a probability space. Any measurable function on a probability space is a random variable. So on $([0,1], \mathcal{B}([0,1]), \lambda)$, $X(\omega) = \omega^2$ is a random variable.*

(m) Formula for transformation of random variables with an example.

Theorem 0.29 *Let X be an n-dimensional absolutely continuous random variable and f_X its density function. Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be a measurable function such that*

(i) g is injective.

(ii) the partial derivatives $\frac{\partial g_i}{\partial x_j} : \mathbb{R}^n \to \mathbb{R}$ exist and are continuous a.e. for all $i, j = 1, \ldots, n$.

(iii) The determinant of the Jacobian matrix of $g^{-1}(y)$ is not zero for a.a. y , i.e. $\det Dg^{-1}(y) \neq 0$ for a.a. $y \in \mathbb{R}^n$.

Then, $Y = g(X)$ is an absolutely continuous n-dimensional random variable with density function given by

$$f_Y(y) = f_X(g^{-1}(y))|\det Dg^{-1}(y)|, \quad y \in \mathbb{R}^n,$$

where $|Dg^{-1}(y)|$ is the determinant of the Jacobian matrix of g^{-1} at the point y.

Example 0.30 *Let X be an exponentially distributed random variable with parameter $\theta > 0$. Then $Y = \log X$ has density:*

$$f_Y(y) = f_X(g^{-1}(y))|(g^{-1})'(y)|,$$

where $y = g(x) = \log(x)$ so $g^{-1}(y) = e^y$ and $|(g^{-1})'(y)| = e^y$ and $y \in (-\infty, \infty)$. We know that $f_X(x) = \lambda e^{-\lambda x}$, $x \in (0, \infty)$. Finally,

$$f_Y(y) = f_X(g^{-1}(y))|(g^{-1})'(y)| = \lambda e^{-\lambda e^y} e^y, \quad y \in \mathbb{R}.$$

This density function gives rise to the distribution function

$$F_Y(y) = e^{-\lambda e^y}, \quad y \in \mathbb{R}$$

which is known as Gumbel distribution.

(n) Expectation of a random variable.

It is simply the integral of the random variable w.r.t. the probability measure of the probability space it is defined on.

Definition 0.31 (Expectation) *Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a random variable X, we define the* expectation *of X, denoted by $\mathbb{E}[X]$, as*

$$\mathbb{E}[X] \triangleq \int_\Omega X(\omega) d\mathbb{P}[\omega].$$

(o) Formula for the expectation with respect to the distribution function.

We already gave it as example in item (k), but here it is again:

Theorem 0.32 (Expectation of a random variable) *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a random variable on this space. Let F_X be the distribution function of X and $g : \mathbb{R} \to \mathbb{R}$ a measurable transformation. Then,*

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) dF_X(x).$$

In particular, if $F_X(x)$ admits an a.e. derivative that is non-identically to zero, then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

Example 0.33 *$([0,1], \mathcal{B}[0,1], \lambda)$ and $X(\omega) = e^{\omega}$. Then*

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{[0,1]} e^{\omega} d\lambda(\omega) = \int_0^1 e^{\omega} d\omega = e - 1.$$

But also via defining its distribution function: for $x \in [1, e]$,

$$F_X(x) = \mathbb{P}[X \leq x] = \mathbb{P}[e^{\omega} \leq x] = \mathbb{P}[\omega \leq \log x] = \log x.$$

Other wise $\mathbb{P}[X \leq x] = 0$ for $x < 1$ and $\mathbb{P}[X \leq x] = 1$ for $x > e$. Then, its density function is given by $f_X(x) = F_X'(x) = \frac{1}{x}$ for $x \in [1, e]$. So,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x dF_X(x) = \int_{\mathbb{R}} x f_X(x) dx = \int_1^e x \frac{1}{x} dx = e - 1.$$

(p) Concept of independent random variables.

Definition 0.34 (Independence of random variables) *A family of (real-valued) random variables $\{X_i\}_{i \in I}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ are independent when the family of sub-σ-algebras $\{X^{-1}(\mathcal{B}(\mathbb{R}))\}_{i \in I} \subset \mathcal{A}$, generated by the random variables $\{X_i\}_{i \in I}$ is an independent family of sub-σ-algebras.*

A more applicable equivalent definition for the bivariate case is the following:

Corollary 0.35 (Independence of random variables) *Given two random variables X, Y in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we say that X and Y are independent if, and only if,*

$$\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[X \leq x]\mathbb{P}[Y \leq y], \quad \text{for every } x, y \in \mathbb{R}.$$

Example 0.36 *The bivariate random vector (X, Y) with distribution*

$$F_{(X,Y)}(x, y) = \mathbb{P}[X \leq x, Y \leq y] = 1 - e^{-(x+y)}, \quad (x, y) \in [0, \infty)^2$$

does not have independent components: i.e. X and Y are not independent because the distribution function $F_{(X,Y)}$ can not be expressed as the product of two functions where each only depend on x and y, respectively.

The bivariate random vector (X, Y) with distribution

$$F_{(X,Y)}(x, y) = \mathbb{P}[X \leq x, Y \leq y] = 1 - e^{-x} - e^{-y} + e^{-(x+y)}, \quad (x, y) \in [0, \infty)^2$$

does have independent components: i.e. X and Y are independent because the distribution function $F_{(X,Y)}$ can not be expressed as the product of two functions where each only depend on x and y, respectively. Indeed,

$$F_{(X,Y)}(x, y) = \mathbb{P}[X \leq x, Y \leq y] = (1 - e^{-x})(1 - e^{-y}), \quad (x, y) \in [0, \infty)^2.$$

Problem 2

Let $(\Omega, \mathcal{A}, \mathbb{P})$ the probability space with $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ and $\mathbb{P}$ is the probability measure defined as

$$\mathbb{P}[B] = \int_B \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, \quad B \in \mathcal{B}(\mathbb{R}),$$

where $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$ are two fixed parameters.

Suppose that the outcomes ω are states of an economy, being negative values of ω seen as *negative states* and positive values as *positive states*.

You run a business in this economy that invests some initial capital C_0. This business has an aleatory rate of return, denoted by ρ. We assume that $\rho : \Omega \to \mathbb{R}$ is a mapping that interconnects the states of the economy with the rate of return of your investment. Your own analysis on the business indicated that $\rho(\omega) = \mu + \sigma\omega$, $\omega \in \Omega$ for fixed parameters $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$.

(a) Prove that $\mathbb{P}$ is indeed a probability measure on the measurable space $(\Omega, \mathcal{A})$.
We need to check three conditions: $\mathbb{P}[\emptyset] = 0$, $\mathbb{P}[\Omega] = 1$ and that the probability of the countable union of disjoint events is the countable sum of probabilities.

Obviously, $\mathbb{P}[\emptyset] = 0$. Also,

$$\mathbb{P}[\Omega] = \int_\Omega \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1.$$

Now, take a countable sequence $\{A_n\}_{n=1}^{\infty}$ of disjoint events. Then

$$\begin{aligned}\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right] &= \int_{\bigcup_{n=1}^{\infty} A_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \int_{\mathbb{R}} \mathbb{I}_{\{\bigcup_{n=1}^{\infty} A_n\}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.\end{aligned}$$

Since the events are disjoint we have

$$\mathbb{I}_{\{\bigcup_{n=1}^{\infty} A_n\}} = \sum_{n=1}^{\infty} \mathbb{I}_{A_n}.$$

It is not immediately clear that we can take out the sum since it is an infinite sum. Let us proceed carefully:

$$\begin{aligned}\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right] &= \int_{\mathbb{R}} \sum_{n=1}^{\infty} \mathbb{I}_{A_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \int_{\mathbb{R}} \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{I}_{A_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.\end{aligned}$$

Now,

$$\left|\sum_{n=1}^{N} \mathbb{I}_{A_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\right| \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}},$$

since $\sum_{n=1}^{N} \mathbb{I}_{A_n} \leq 1$ insomuch that A_n are disjoint. Since the sequence of functions $\{h_N\}_N$ $h_N(z) = \sum_{n=1}^{N} \mathbb{I}_{A_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ are bounded by an integrable function, by Lebesgue's dominated convergence theorem, we can move the limit outside the integral and then we can also move the sum outside the integral since it is a finite sum:

$$\begin{aligned}\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right] &= \int_{\mathbb{R}} \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{I}_{A_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \int_{\mathbb{R}} \mathbb{I}_{A_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \sum_{n=1}^{\infty} \int_{\mathbb{R}} \mathbb{I}_{A_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \sum_{n=1}^{\infty} \int_{A_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \mathbb{P}[A_n].\end{aligned}$$

(b) Prove that ρ is a random variable.

The mapping $\rho : \mathbb{R} \to \mathbb{R}$ as $\rho(\omega) = \mu + \sigma\omega$ is a random variable since it maps Borel sets from $\mathbb{R}$ in the image space to Borel sets in $\mathbb{R}$ in the domain. We just need to check that

$$\{\omega \in \mathbb{R} : \rho(\omega) \leq x\}$$

is a Borel set for all x. Indeed,

$$\{\omega \in \mathbb{R} : \rho(\omega) \leq x\} = \left\{\omega \in \mathbb{R} : \omega \leq \frac{x-\mu}{\sigma}\right\} = \left(-\infty, \frac{x-\mu}{\sigma}\right],$$

which is obviously a Borel set.

Assume that this investment runs for t years and that returns are compounded continuously. This essentially means that after t years, your capital will have increased (or decreased) to

$$C(\omega) = C_0 e^{\rho(\omega)t}, \quad \omega \in \Omega.$$

Note that $C \in (0, \infty)$.

(c) Prove that C, the capital gains from investing C_0 is indeed a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$.

Note that if we have $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to$ such that Dom $g \subseteq$ Im f and f, g are measurable then $f \circ g$ are measurable: Indeed, take a Borel set B, then

$$(f \circ g)^{-1}(B) = g^{-1} \circ f^{-1}(B).$$

Since f is measurable then $f^{-1}(B) = \widetilde{B}$ is a Borel set, so

$$(f \circ g)^{-1}(B) = g^{-1}(\widetilde{B}).$$

Since g is measurable, then $g^{-1}(\widetilde{B})$ is a Borel set, hence $f \circ g$ is measurable.

Now, $\{C_0 e^{\omega} \leq xt\} = \{\omega \leq \log \frac{xt}{C_0}\} = (-\infty, \log \frac{xt}{C_0}]$, $x \in (0, \infty)$ and $\{C_0 e^{\omega} \leq xt\} = \emptyset$ if $x \in (-\infty, 0]$ which are both Borel sets. So the exponential function multiplied by a constant is measurable. Now, the image of ρ is the domain of $C_0 e^{\omega}$ so their composition $C = C_0 e^{t\rho}$ is measurable.

You could also have proved this directly for C.

(d) Find the law of C and compute the expected gains, i.e. $\mathbb{E}[C]$.

The law of C is

$$\mu_C(B) = \mathbb{P}[C \in B], \quad B \in \mathcal{B}(\mathbb{R}).$$

Now, let us restrict our choice to a Borel set of the form $B = [a, b]$, $a < b$. Then

$$\begin{aligned}
\mu_C(B) &= \mathbb{P}[C \in B] \\
&= \mathbb{P}[C_0 e^{t\rho} \in [a, b]] \\
&= \mathbb{P}[a \leq C_0 e^{\mu t + \sigma t \omega} \leq b] \\
&= \mathbb{P}\left[\log \frac{a}{C_0} \leq \mu t + \sigma t \omega \leq \log \frac{b}{C_0}\right] \\
&= \mathbb{P}\left[\frac{\log \frac{a}{C_0} - \mu t}{\sigma t} \leq \omega \leq \frac{\log \frac{b}{C_0} - \mu t}{\sigma t}\right] \\
&= \int_{[\widetilde{a}(a), \widetilde{b}(b)]} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz,
\end{aligned}$$

where

$$[\widetilde{a}(a), \widetilde{b}(b)] = \left[\frac{\log \frac{a}{C_0} - \mu t}{\sigma t}, \frac{\log \frac{b}{C_0} - \mu t}{\sigma t}\right].$$

Now, the distribution function is the law applied to the Borel set $B = (-\infty, x]$, this means taking $a = -\infty$ and $b = x$ above. Thus,

$$F_C(x) = \int_{[\widetilde{a}(-\infty), \widetilde{b}(x)]} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{-\infty}^{\widetilde{b}(x)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.$$

Then, the density function of C is given by the derivative of F_C. Using Leibniz' rule (chain rule) we have

$$f_C(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\widetilde{b}(x)^2} \widetilde{b}'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\widetilde{b}(x)^2} \frac{1}{\sigma t x}, \quad x \in \mathbb{R}.$$

The expectation of C using the *law of the unconscious statistician* is

$$\mathbb{E}[C] = \int_{-\infty}^{\infty} x f_C(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 t^2}} e^{-\frac{1}{2\sigma^2 t^2}(\log x - \widetilde{\mu})} dx,$$

where $\widetilde{\mu} = \log C_0 + \mu t$.

We can see that the above integral corresponds to the expectation of a log-normal distribution with parameters $\widetilde{\mu}$ and $\widetilde{\sigma}^2 = \sigma^2 t^2$. You may compute it by using a change of variables of the form $y = \log x$. The result is

$$\mathbb{E}[C] = C_0 e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

(e) You are only interested in running this business if your gains C exceed a given threshold $K > 0$, i.e. $C > K$. Find the distribution of the random variable $Y = (C-K)\mathbb{I}_{\{C>K\}}$ and the expectation of this new random variable Y. What type of random variable is Y? Note: you will not be able to find an analytical closed expression, but you can leave the expression in terms of integrals.

Observe that the random variable Y can take the values 0 (if $C \leq K$ or any positive number on $(0, \infty)$ if $C > K$. So, the random variable Y takes values in $[0, \infty)$ where 0 is an atom, i.e. $\mathbb{P}[Y = 0] > 0$. This latter fact implies that the distribution function of Y will be discontinuous at $x = 0$.

$$\mathbb{P}[Y = 0] = \mathbb{P}[C \leq K] = \int_{-\infty}^{K} h(z)dz = F_C(K),$$

where F_C is the distribution function of C. Now if $y > 0$ we have

$$\begin{aligned}
\mathbb{P}[Y \leq y] &= \mathbb{P}[(C-K)\mathbb{I}_{\{C>K\}} \leq y] \\
&= \underbrace{\mathbb{P}[(C-K)\mathbb{I}_{\{C>K\}} \leq y | C \leq K]}_{=1} \mathbb{P}[C \leq K] \\
&\quad + \mathbb{P}[(C-K)\mathbb{I}_{\{C>K\}} \leq y | C > K]\mathbb{P}[C > K] \\
&= \mathbb{P}[C \leq K] + \mathbb{P}[C \leq y + K | C > K](1 - F_C(K)) \\
&= F_C(K) + \frac{\mathbb{P}[K < C \leq y + K]}{1 - F_C(K)}(1 - F_C(K)) \\
&= F_C(K) + F_C(y + K) - F_C(K) \\
&= F_C(y + K).
\end{aligned}$$

Hence,

$$F_Y(y) = F_C(K)\mathbb{I}_{\{y=0\}} + F_C(y + K)\mathbb{I}_{\{y>0\}}, \quad y \in [0, \infty).$$

Observe that F_Y has a jump at $y = 0$. Since $\lim_{\substack{y \to 0 \\ y<0}} F_Y(y) = 0$ and $\lim_{\substack{y \to 0 \\ y>0}} F_Y(y) = F_C(K) \neq 0$. But F_Y is indeed right-continuous. Recall that if F is a function

with a discontinuity at $x = a$ and otherwise differentiable with derivative f, then

$$\int_{\mathbb{R}} g(x)dF(x) = \int_{-\infty}^{\infty} g(x)f(x)dx + g(a)(F(a) - F(a-)),$$

where $F(a-)$ is the left limit of F at the point a. Since F_Y has a discontinuity at 0 and is otherwise differentiable we have

$$\mathbb{E}[Y] = \int_{\mathbb{R}} y dF_Y(y) = \int_{-\infty}^{\infty} y f_Y(y)dy + 0 \cdot (F_Y(0) - F_Y(0-)) = \int_{-\infty}^{\infty} y f_Y(y)dy.$$

Now, $f_Y(y) = f_C(y + K)$, $y > 0$. Thus,

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_C(y+K)dy = \int_{K}^{\infty} (y-K) f_C(y)dy = \int_{K}^{\infty} z f_C(z)dz - K(1-F_C(K)).$$

Observe that we could have reached the same conclusion by simply using the law of unconscious statistician: $\mathbb{E}[g(X)] = \int g(x)f_X(x)dx$. So,

$$\mathbb{E}[(C - K)\mathbb{I}_{\{C>K\}}] = \int_{K}^{\infty} (z - K) f_C(z)dz.$$

The formula for $\mathbb{E}[Y]$ is actually known as (a specific version of) Black-Scholes formula.

Problem 3 (Hölder's inequality)

The proof of this exercise can be found in many places in the internet, but try to do it yourself and understand the proof.

(a) **Young's inequality**: Let $a, b \in [0, \infty)$ and $p, q \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Prove that

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Method I: If $a = b = 0$ then the inequality is trivially satisfied. Assume $a > 0$ and $b > 0$. Set $t = \frac{1}{p}$ then it follows that $1 - t = \frac{1}{q}$ since $\frac{1}{p} = 1 - \frac{1}{q}$ by assumption. Then,

$$\frac{a^p}{p} + \frac{b^q}{q} = ta^p + (1 - t)b^q.$$

Recall the definition of a concave function: a function f is said to be concave if for every $t \in [0, 1]$ an any two points x, y, $x \neq y$ one has

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y).$$

The above inequality has the geometric intuition that given two points x, y, $x \neq y$, the cord joining the points $(x, f(x))$ and $(y, f(y))$ remains below the graph of f.

Now, since the logarithm function is concave, and we have $t \in (0,1)$ then we know that

$$\log(ta^p+(1-t)b^q) \leq t\log(a^p)+(1-t)\log(b^q) = tp\log(a)+(1-t)q\log(b) = \log(ab),$$

where we used that $tp = 1$ since $t = \frac{1}{p}$ and that $\log(a) + \log(b) = \log(ab)$.

Then, Young's inequality follows by exponentiating both sides of the above inequality.

Method II: One can of course use standard one-variable calculus as well to prove Young's inequality by showing that a function is increasing. The proof can be found here: `https://proofwiki.org/wiki/Young%27s_Inequality_for_Products#google_vignette`

(b) **Hölder's inequality**: Let $p, q \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$ and $f, g : \mathbb{R} \to \mathbb{R}$ two integrable functions. Show that

$$\int |fg| \leq \left(\int |f|^p\right)^{\frac{1}{p}} \left(\int |g|^q\right)^{\frac{1}{q}}.$$

Hint: Use Young's inequality.

If $\int |f| = 0$ or $\int |g| = 0$ then the inequality is true. Assume $\int |f| > 0$ and $\int |g| > 0$. Define the functions

$$\widetilde{f} = \frac{f}{C_p}, \quad \widetilde{g} = \frac{g}{C_q},$$

where the constants C_p and C_q are, respectively, defined by

$$C_p \triangleq \left(\int |f|^p\right)^{\frac{1}{p}} \quad C_q \triangleq \left(\int |g|^q\right)^{\frac{1}{q}}.$$

Then we can apply Young's inequality to the real numbers $a = \widetilde{f}(x)$ and $b = \widetilde{g}(x)$, $x \in \mathbb{R}$. Thus,

$$|\widetilde{f}(x)\widetilde{g}(x)| \leq \frac{|\widetilde{f}(x)|^p}{p} + \frac{|\widetilde{g}(x)|^q}{q}.$$

Then, we have

$$\frac{|f(x)g(x)|}{C_pC_q} \leq \frac{1}{C_p^p}\frac{|f(x)|^p}{p} + \frac{1}{C_q^q}\frac{|g(x)|^q}{q}.$$

Furthermore,

$$|f(x)g(x)| \leq C_pC_q\left(\frac{1}{C_p^p}\frac{|f(x)|^p}{p} + \frac{1}{C_q^q}\frac{|g(x)|^q}{q}\right).$$

Integrating over x we get

$$\int |f(x)g(x)| \leq C_pC_q\left(\frac{1}{C_p^p}\frac{\int |f(x)|^p}{p} + \frac{1}{C_q^q}\frac{\int |g(x)|^q}{q}\right).$$

But note that

$$\frac{1}{C_p^p}\frac{\int |f(x)|^p}{p} + \frac{1}{C_q^q}\frac{\int |g(x)|^q}{q} = \frac{1}{p} + \frac{1}{q} = 1.$$

Hence,

$$\int |f(x)g(x)| \leq C_pC_q,$$

which is precisely Hölder's inequality.

(c) Let X, Y two random variables and p, q as in the previous items. Prove that

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}}\mathbb{E}[|Y|^q]^{\frac{1}{q}}.$$

Hölder's inequality in item (b) yields for general measures. Then this is a particular case for integrals w.r.t. $\mathbb{P}$,

$$\int_\Omega |X(\omega)Y(\omega)|d\mathbb{P}(\omega) \leq \left(\int_\Omega |X(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}}\left(\int_\Omega |Y(\omega)|^q d\mathbb{P}(\omega)\right)^{\frac{1}{q}}$$

which is exactly

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}}\mathbb{E}[|Y|^q]^{\frac{1}{q}}.$$