Tools of Modern Probability

Imre Péter Tóth

Practice exercises

Contents

1 Gaussian integrals

1.1 Find all continuous functions $f : \mathbb{R}^2 \to \mathbb{R}$ that are rotation invariant and also of product form. That is, there are functions $g : [0,\infty) \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ such that, for every $x, y \in \mathbb{R}$

$$f(x,y) = g(\sqrt{x^2+y^2}) = h(x)h(y).$$

(Hint: write everything as the function of the ***square*** *of the radius, e.g. by defining* $u := x^2$*,* $v := y^2$ *and* $G(z) := g(\sqrt{z})$*. Then you should get* $G(u+v) = constG(u)G(v)$*. Now study the logarithm of* G*.)*

1.2 Use the integral substitution $\frac{y^2}{2} := a(x-m)^2$ to show that

$$\int_{-\infty}^{\infty} e^{-a(x-m)^2}\,\mathrm{d}x = \sqrt{\frac{\pi}{a}} \tag{1}$$

whenever $m \in \mathbb{R}$ and $0 < a \in \mathbb{R}$. We know form class that the value of the integral is $\sqrt{2\pi}$ when $m = 0$ and $a = \frac{1}{2}$.

1.3 Let $f(x_1,\dots,x_d) = e^{-\frac{x_1^2+\dots+x_d^2}{2}}$, and let $V = \int_{\mathbb{R}^d} f(\underline{x})\,\mathrm{d}\underline{x}$.

- Calculate V using that f is a product:

$$f(x_1,\dots,x_d) = e^{-\frac{x_1^2}{2}} \cdot e^{-\frac{x_2^2}{2}} \cdot \dots \cdot e^{-\frac{x_d^2}{2}}.$$

- Write V as a one-dimensional integral using polar coordinate substitution.

- Compare the two results to get that

$$c_d = \frac{\sqrt{2\pi}^d}{\int_0^\infty r^{d-1} e^{-\frac{r^2}{2}}\,\mathrm{d}r}.$$

1.4 Calculate $A_n := \int_0^{\frac{\pi}{2}} \cos^n x \,\mathrm{d}x$ for every $n = 0, 1, 2, \ldots$ the hard way: if $n \geq 2$, then

$$A_n = \int_0^{\frac{\pi}{2}} (1 - \sin^2 x) \cos^{n-2} x \,\mathrm{d}x = A_{n-2} - \int_0^{\frac{\pi}{2}} [\sin x] \left[\sin x \cos^{n-2} x\right] \,\mathrm{d}x,$$

and you can use integration by parts in the second term.

1.5 Let $B_d \subset \mathbb{R}^d$ be the unit ball in R^d meaning

$$B_d := \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d \middle| x_1^2 + \cdots + x_d^2 \leq 1\right\}.$$

(Compare the definition of the sphere – note the inequality here.) Let b_d be the d-dimensional volume of B_d. Calculate b_d.

(Hint: the volume is the integral of the indicator function. Use the theorem about polar coordinate substitution in d dimensions.)

1.6 Try to calculate b_d of the previous exercise the hard way: slice the $d+1$-dimensional sphere into d-dimensional ones to see that

$$b_{d+1} = \int_{-1}^{1} b_d \sqrt{1 - x^2}^d \,\mathrm{d}x.$$

2 Euler gamma function

2.1 For $s > 0$ let

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} \,\mathrm{d}x$$

be the Euler gamma function. Check that $\Gamma(s+1) = s\Gamma(s)$ for all $s > 0$. Check by induction that $\Gamma(n+1) = n!$ for all $n \in \mathbb{N}$.

2.2 Calculate $\Gamma\left(\frac{1}{2}\right)$. Express $\Gamma(s)$ for every half-integer $s > 0$ using factorials.

2.3 Fix some $s, t > 0$. Consider $f : (0, \infty) \times (0, \infty) \to \mathbb{R}$ defined by $f(x, y) := x^{s-1} e^{-x} y^{t-1} e^{-y}$ (for all $x, y > 0$). Calculate $\int_{(0,\infty)^2} f(x, y) \,\mathrm{d}x \,\mathrm{d}y$ in two different ways:

a.) By using that f has product form,

b.) using the substitution $u := x + y$, $\xi := \frac{y}{x+y}$. *(If it's easier, you can do this in two steps: first $u := x + y$, $v := y$; second $\xi := v/u$.)*

Comparing the two results, express the Beta function $B(s, t) := \int_0^1 (1 - \xi)^{s-1} \xi^{t-1} \,\mathrm{d}\xi$ using the Euler gamma function.

2.4 Calculate $A_n := \int_0^{\frac{\pi}{2}} \cos^n x \,\mathrm{d}x$ for every $n = 0, 1, 2, \ldots$ using the substitution $\xi := \cos x$ and the result of the previous exercise.

3 Almost Gaussian integrals

3.1 Describe the asymptotic behaviour of the integral $I_n := \int_{-1}^{1} \sqrt{1-x^2}^n \, \mathrm{d}x$ as $n \to \infty$.

3.2 Describe the asymptotic behaviour of the integral $I_n := \int_{-2}^{2} \sqrt{4-x^2}^n \, \mathrm{d}x$ as $n \to \infty$.

3.3 Let

$$f_n(x) = \begin{cases} \cos^n x & \text{if } x \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ 0 & \text{if not} \end{cases}.$$

Let $g_n(x) = f_n(v_n x)$, where the scaling factors v_n are chosen appropriately, so that $\int_{\mathbb{R}} g_n \to 1$ *(More precisely: g_n should be integrated on all of its domain.)* Find the limit $g(x) := lim_{n\to\infty} g_n(x)$.

3.4 Let $f_n(x) = \sqrt{4-x^2}^n$ (for $x \in [-2, 2]$), and let $g_n(x) = u_n f_n(v_n x)$, where the scaling factors u_n and v_n are chosen appropriately, so that $g_n(0) \to 1$ and $\int_{\mathbb{R}} g_n \to 1$ *(More precisely: g_n should be integrated on all of its domain.)* Find the limit $g(x) := lim_{n\to\infty} g_n(x)$.

3.5 Let $a < 0 < b$ and let $h : [a, b] \to \mathbb{R}$ be twice differentiable with a unique non-degenerate local maximum at 0. Denote $A := h(0)$ and $B := -h''(0)$. Let $f_n : [a, b] \to \mathbb{R}$ with $f_n(x) = e^{nh(x)}$. Now let $u_n > 0$ and $v_n > 0$ be two sequences of scaling factors, and define g_n as

$$g_n(x) := u_n f_n(v_n x),$$

for the $x \in \mathbb{R}$ where this makes sense. (This means stretching the graph of f_n vertically with a factor u_n and shrinking it horizontally with a factor v_n.)

a.) How should we choose u_n to make sure that $g_n(0) \to 0$ as $n \to \infty$? *(Of course, there are many such sequences: if u_n works and $\bar{u}_n \sim u_n$, then $\bar{u}_n$ works as well. So give a simple example.)*

b.) Fix u_n as in the previous part. Now how should be choose v_n to make sure that

$$\int_{D_n} g_n(x) \, \mathrm{d}x \to 1$$

as $n \to \infty$? (Here let D_n denote the domain of g_n.)

c.) With u_n and v_n chosen as above, calculate $g(x) := \lim_{n\to\infty} g_n(x)$ for all $x \in \mathbb{R}$.

4 Stirling's approximation

4.1 Let the random vector $V = (V_1, \ldots, V_n) \in \mathbb{R}^n$ be uniformly distributed on the (surface of the) $(n-1)$-dimensional sphere of radius $\sqrt{2nE}$ in $\mathbb{R}^n$. Let f_n denote the density of the first marginal V_1 (which is itself a random variable in $\mathbb{R}$, and, of course, its density depends on n). Calculate $f_n(x)$ for every n. Find the limit $f(x) := \lim_{n\to\infty} f_n(x)$.

4.2 *[DeMoivre-Laplace Central Limit Theorem]* We toss a biased coin (where the probability of "heads" is some $p \in (0, 1)$) n times independently. Let $q = 1 - p$. Let X be the number of heads we see. So X is binomially distributed with parameters n and p, meaning

$$\mathbb{P}(X = k) = Bin(k; n, p) := \binom{n}{k} p^k q^{n-k} \quad \text{for } k = 0, 1, \ldots, n.$$

It is known that X has expectation $\mathbb{E}X = np$ and standard deviation $DX = \sqrt{VarX} = \sqrt{npq}$, so let $Y := \frac{X-np}{\sqrt{npq}}$ be the normalized version of X (which now has expectation 0 and standard deviation 1). Of course, Y is still a discrete random variable, taking only values from a grid of points which are $\frac{1}{\sqrt{npq}}$ apart.

Let us fix $x \in \mathbb{R}$, and choose $k \in \mathbb{Z}$ such that $x \approx \frac{k-np}{\sqrt{npq}}$ as closely as possible, so k is $np + x\sqrt{npq}$ rounded to the nearest integer. Let

$$f_n(x) := \frac{\mathbb{P}(Y = \frac{k-np}{\sqrt{npq}})}{\frac{1}{\sqrt{npq}}} = \sqrt{npq}\mathbb{P}(X = k)$$

be the logical guess for an "approximate density" of Y at x.

Calculate the limit $f(x) := \lim_{n\to\infty} f_n(x)$.

Hint:

Use Stirling's approximation $n! \sim \frac{n^n\sqrt{2\pi n}}{e^n}$, and the fact that $k = np + x\sqrt{npq} + \Delta$, where $\Delta = \Delta(n, x) \in [-\frac{1}{2}, \frac{1}{2}]$, so $\Delta = O(1)$. Use this in the following forms:

$$k = np + x\sqrt{npq} + \Delta \quad , \quad n - k = nq - x\sqrt{npq} - \Delta \tag{2}$$

$$\frac{k}{np} = 1 + x\sqrt{\frac{q}{np}} + \frac{\Delta}{np} \quad , \quad \frac{n-k}{nq} = 1 - x\sqrt{\frac{p}{nq}} - \frac{\Delta}{nq} \tag{3}$$

$$\frac{k}{np} = 1 + o(1) \quad , \quad \frac{n-k}{nq} = 1 + o(1) \tag{4}$$

Notice that (2) is a bit stronger than if we only wrote $k = np + x\sqrt{npq} + O(1)$ and $n - k = nq - x\sqrt{npq} + O(1)$. This will be important, since Δ will cancel out at some point.

At some point the calculation may become more transparent if you calculate the logarithm of $f_n(x)$.

5 Basics of measure theory

5.1 Define a σ-algebra as follows:

Definition 1 *For a nonempty set Ω, a family $\mathcal{F}$ of subsets of ω (i.e. $\mathcal{F} \subset 2^\Omega$, where $2^\Omega := \{A : A \subset \Omega\}$ is the power set of Ω) is called a σ-algebra over Ω if*

- $\emptyset \in \mathcal{F}$
- *if $A \in \mathcal{F}$, then $A^C := \Omega \setminus A \in \mathcal{F}$ (that is, $\mathcal{F}$ is closed under complement taking)*
- *if $A_1, A_2, \dots \in \mathcal{F}$, then $(\cup_{i=1}^\infty A_i) \in \mathcal{F}$ (that is, $\mathcal{F}$ is closed under countable union).*

Show from this definition that a σ-algebra is closed under countable intersection, and under finite union and intersection.

5.2 *Continuity of the measure*

(a) Prove the following:

Theorem 1 *(Continuity of the measure)*

i. If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $A_1, A_2, \dots$ is an increasing sequence of measurable sets (i.e. $A_i \in \mathcal{F}$ and $A_i \subset A_{i+1}$ for all i), then $\mu(\cup_{i=1}^\infty A_i) = \lim_{i\to\infty} \mu(A_i)$ (and both sides of the equation make sense).

ii. If $(\Omega, \mathcal{F}, \mu)$ is a measure space, $A_1, A_2, \dots$ is a decreasing sequence of measurable sets (i.e. $A_i \in \mathcal{F}$ and $A_i \supset A_{i+1}$ for all i) and $\mu(A_1) < \infty$, then $\mu(\cap_{i=1}^\infty A_i) = \lim_{i\to\infty} \mu(A_i)$ (and both sides of the equation make sense).

(b) Show that in the second statement the condition $\mu(A_1) < \infty$ is needed, by constructing a counterexample for the statement when this condition does not hold.

5.3 (a) We toss a biased coin, on which the probability of heads is some $0 \le p \le 1$. Define the random variable ξ as the indicator function of tossing heads, that is

$$\xi := \begin{cases} 0, & \text{if tails} \\ 1, & \text{if heads} \end{cases}.$$

i. Describe the distribution of ξ (called the Bernoulli distribution with parameter p) in the "classical" way, listing possible values and their probabilities,

ii. and also by describing the distribution as a measure on $\mathbb{R}$, giving the weight $\mathbb{P}(\xi \in B)$ of every Borel subset B of $\mathbb{R}$.

iii. Calculate the expectation of ξ.

(b) We toss the previous biased coin n times, and denote by X the *number of heads* tossed.

i. Describe the distribution of X (called the Binomial distribution with parameters (n, p)) by listing possible values and their probabilities.

ii. Calculate the expectation of X by integration (actually summation in this case) using its distribution,

iii. and also by noticing that $X = \xi_1 + \xi_2 + \cdots + \xi_n$, where ξ_i is the indicator of the i-th toss being heads, and using linearity of the expectation.

5.4 The *ternary* number $0.a_1a_2a_3\ldots$ is the analogue of the usual decimal fraction, but writing numbers in base 3. That is, for any sequence $a_1, a_2, a_3, \ldots$ with $a_n \in \{0, 1, 2\}$, by definition

$$0.a_1a_2a_3\cdots := \sum_{n=1}^{\infty} \frac{a_n}{3^n}.$$

Now let us construct the ternary fraction form of a random real number X via a sequence of fair coin tosses, such that we rule out the digit 1. That is,

$$a_n := \begin{cases} 0, & \text{if the } n\text{-th toss is tails,} \\ 2, & \text{if the } n\text{-th toss is heads} \end{cases},$$

and setting $X = 0.a_1a_2a_3\ldots$ (ternary). In this way, X is a "uniformly" chosen random point of the famous *middle-third Cantor set* C defined as

$$C := \left\{ \sum_{n=1}^{\infty} \frac{a_n}{3^n}, \; a_n \in \{0, 2\} \; (n = 1, 2, \ldots) \right\}.$$

Show that

(a) The distribution of X gives zero weight to every point – that is, $\mathbb{P}(X = x) = 0$ for every $x \in \mathbb{R}$. (As a consequence, the cumulative distribution function of X is continuous.)

(b) The distribution of X is not absolutely continuous w.r.t the Lebesgue measure on $\mathbb{R}$.

5.5 Let V be a random vector in $\mathbb{R}^n$ with an n-dimensional standard Gaussian distribution, meaning that it has density

$$f(v_1, \ldots, v_n) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{v_1^2 + \cdots + v_n^2}{2}}.$$

Think of V as the velocity vector of a particle with mass m, so the energy is $E = \frac{m}{2}V^2$. Calculate the distribution of the random variable E. (Meaning: calculate the distribution function and/or the density, and tell the name of the distribution.)

5.6 *Usefulness of the linearity of the expectation.* A building has 10 floors, not including the ground floor. On the ground floor, 10 people get into the elevator, and every one of them chooses a destination at random, uniformly out of the 10 floors, independently of the others. Let X denote the number of floors *on which the elevator stops* – i.e. the number of floors that were chosen by at least one person. Calculate the expectation and the variance of X. *(hint: First notice that the distribution of X is hard to calculate. Find a way to calculate the expectation and the variance without that.)*

5.7 Let $X = [0, 1]$ and let μ be Lebesgue measure on X. Let $f(x) = x^2$. Describe the measure $f_*\mu$

a.) by calculating $(f_*\mu)([a, b])$ for every interval $[a, b] \subset \mathbb{R}$

b.) by giving the density of $f_*\mu$ with respect to Lebesgue measure.

5.8 Let $X = \{(a_1, a_2, \dots) \mid a_k \in \{0, 1\}$ for every $k\}$ be the set of $\{0, 1\}$-sequences. Let μ be the measure on X for which

$$\mu(\{(a_1, a_2, \dots) \in X \mid a_1 = b_1, \dots, a_N = b_N\}) = \frac{1}{2^N}$$

for every $b_1, \dots, b_N \in \{0, 1\}$. Let $f : X \to \mathbb{R}$ be defined as

$$f(a_1, a_2, \dots) := \sum_{k=1}^{\infty} \frac{a_k}{2^k}.$$

Describe the measure $f_*\mu$

a.) by calculating $(f_*\mu)([a, b])$ for every interval $[a, b] \subset \mathbb{R}$

b.) by giving the density of $f_*\mu$ with respect to Lebesgue measure.

5.9 Let λ be Lebesgue measure and χ be counting measure on $\mathbb{R}$ (with the Borel σ-algebra). Show that λ does not have a density with respect to χ. (Hint: consider 1-element sets.)

5.10 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A \in \mathcal{F}$. Define $X : \Omega \to \mathbb{R}$ as $X(\omega) = \mathbf{1}_A(\omega)$ and let $\mu = X_*\mathbb{P}$ be the distribution of X. Show that μ is absolutely continuous w.r.t counting measure, show that it also has a density. What is the density?

5.11 Let X be a discrete random variable and let μ be its distribution. Give the density of μ w.r.t. counting measure.

6 Convergence of sequences of functions

6.1 Consider the following measure spaces (X, μ):

I. $X = [0, 1]$, μ is Lebesgue measure.

II. $X = [0, \infty)$, μ is Lebesgue measure.

III. $X = \{1, 2 \dots, N\}$, μ is counting measure.

IV. $X = \{1, 2 \dots\}$, μ is counting measure.

Show examples of functions $f_1, f_2, \dots$ and f from X to $\mathbb{R}$ such that f_n converges to f

a.) almost everywhere, but not in L^1,

b.) in L^1, but not almost everywhere, c.) in L^1, but not in L^2,

d.) in L^2, but not in L^1.

6.2 The characteristic function of a random variable X is the function $\Psi : \mathbb{R} \to \mathbb{C}$ defined as $\Psi(t) = \mathbb{E}(e^{itX})$. Calculate the characteristic function of

(a) The Bernoulli distribution $B(p)$

(b) The "pessimistic geometric distribution with parameter p" – that is, the distribution μ on $\{0, 1, 2 \dots\}$ with weights $\mu(\{k\}) = (1-p)p^k$ $(k = 0, 1, 2 \dots)$.

(c) The "optimistic geometric distribution with parameter p" – that is, the distribution ν on $\{1, 2, 3, \dots\}$ with weights $\nu(\{k\}) = (1-p)p^{k-1}$ $(k = 1, 2 \dots)$.

(d) The Poisson distribution with parameter λ – that is, the distribution η on $\{0, 1, 2 \dots\}$ with weights $\eta(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!}$ $(k = 0, 1, 2 \dots)$.

(e) The exponential distribution with parameter λ – that is, the distribution on $\mathbb{R}$ with density (w.r.t. Lebesgue measure)

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0 \\ 0, & \text{if not} \end{cases} .$$

6.3 For a real values random variable X, the characteristic function of X is $\psi_X : \mathbb{R} \to \mathbb{C}$ defined as $\psi_X(t) := \mathbb{E}\left(e^{itX}\right)$, where $i \in \mathbb{C}$ is the imaginary unit. Show that $\psi_X(t)$ exists for every $t \in \mathbb{R}$.

6.4 For a probability distribution ν on $\mathbb{R}$, the characteristic function of ν is $\psi_\nu : \mathbb{R} \to \mathbb{C}$ defined as $\psi_\nu(t) := \int_{\mathbb{R}} e^{itx}\,\mathrm{d}\nu(x)$, where $i \in \mathbb{C}$ is the imaginary unit. Show that $\psi_\nu(t)$ exists for every $t \in \mathbb{R}$.

6.5 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X : \Omega \to \mathbb{R}$ be a random variable and let $\nu = X_*\mathbb{P}$ be its distribution. Show that $\psi_X = \psi_\nu$, where ψ_X and ψ_μ are the characteristic functions defined in exercises 3 and 4.

6.6 *Dominated convergence and continuous differentiability of the characteristic function.* The Lebesgue dominated convergence theorem is the following

Theorem 2 (dominated convergence) *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_1, f_2, \dots$ measurable real valued functions on Ω which converge to the limit function pointwise, μ-almost everywhere. (That is, $\lim_{n\to\infty} f_n(x) = f(x)$ for every $x \in \Omega$, except possibly for a set of x-es with μ-measure zero.) Assume furthermore that the f_n admit a common* integrable dominating function*: there exists a $g : \Omega \to \mathbb{R}$ such that $|f_n(x)| \le g(x)$ for every $x \in \Omega$ and $n \in \mathbb{N}$, and $\int_\Omega g\,\mathrm{d}\mu < \infty$. Then (all the f_n and also f are integrable and)*

$$\lim_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu.$$

Use this theorem to prove the following:

a.) **Theorem 3 (Continuity of the characteristic function, 1)** *For any real valued random variable X, its characteristic function $\psi_X(t) = \mathbb{E}(e^{itX})$ is continuous.*

b.) **Theorem 4 (Continuity of the characteristic function, 2)** *For any probability distribution ν on $\mathbb{R}$, its characteristic function $\psi_\nu(t) = \int_{\mathbb{R}} e^{itx}\,\mathrm{d}\nu(x)$ is continuous.*

c.) **Theorem 5 (Differentiability of the characteristic function, 1)** *Let X be a real valued random variable, its characteristic function $\psi_X(t) = \mathbb{E}(e^{itX})$. If X is integrable, then ψ_X is differentiable.*

d.) **Theorem 6 (Differentiability of the characteristic function, 2)** *Let ν be a probability distribution on $\mathbb{R}$, its characteristic function $\psi_\nu(t) = \int_{\mathbb{R}} e^{itx}\, \mathrm{d}\nu(x)$. If $\mathbb{E}\nu \in \mathbb{R}$, then ψ_ν is differentiable.*

e.) **Theorem 7 (Continuous differentiability of the characteristic function, 1)** *Let X be a real valued random variable, its characteristic function $\psi_X(t) = \mathbb{E}(e^{itX})$. If X is integrable, then ψ'_X is continuous.*

f.) **Theorem 8 (Continuous differentiability of the characteristic function, 2)** *Let ν be a probability distribution on $\mathbb{R}$, its characteristic function $\psi_\nu(t) = \int_{\mathbb{R}} e^{itx}\, \mathrm{d}\nu(x)$. If $\mathbb{E}\nu \in \mathbb{R}$, then ψ'_ν is continuous.*

6.7 *Exchangeability of integral and limit.* Consider the sequences of functions $f_n : [0,1] \to \mathbb{R}$ and $g_n : [0,1] \to \mathbb{R}$ concerning their pointwise limits and the limits of their integrals. Do there exist integrable functions $f : [0,1] \to \mathbb{R}$ and $g : [0,1] \to \mathbb{R}$, such that $f_n(x) \to f(x)$ and $g_n(x) \to g(x)$ for Lebesgue almost every $x \in [0,1]$? What is $\lim\limits_{n\to\infty}\left(\int\limits_0^1 f_n(x)dx\right)$ and $\lim\limits_{n\to\infty}\left(\int\limits_0^1 g_n(x)dx\right)$? Are the conditions of the dominated and monotone convergence theorems and the Fatou lemma satisfied? If yes, what do these theorems ensure about these specific examples?

(a)

$$f_n(x) = \begin{cases} n^2x & \text{if } 0 \le x < 1/n, \\ 2n - n^2x & \text{if } 1/n \le x \le 2/n, \\ 0 & \text{otherwise.} \end{cases}$$

(b) Write n as $n = 2^k + l$, where $k = 0, 1, 2\dots$ and $l = 0, 1, \dots, 2^k - 1$ (this can be done in a unique way for every n). Now let

$$g_n(x) = \begin{cases} 1 & \text{if } \frac{l}{2^k} \le x < \frac{l+1}{2^k}, \\ 0 & \text{otherwise.} \end{cases}$$

6.8 *Exchangeability of integrals.* Consider the following function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(x) = \begin{cases} 1 & \text{if} \quad 0 < x,\ 0 < y \text{ and } 0 \le x - y \le 1, \\ -1 & \text{if} \quad 0 < x,\ 0 < y \text{ and } 0 < y - x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $\int\limits_{-\infty}^{+\infty}\left(\int\limits_{-\infty}^{+\infty} f(x,y)dx\right)dy$ and $\int\limits_{-\infty}^{+\infty}\left(\int\limits_{-\infty}^{+\infty} f(x,y)dy\right)dx$. What's the situation with the Fubini theorem?

6.9 *Weak convergence and densities.* Prove the following

Theorem 9 *Let $\mu_1, \mu_2, \dots$ and μ be a sequence of probability distributions on $\mathbb{R}$ which are absolutely continouos w.r.t. Lebesgue measure. Denote their densities by $f_1, f_2, \dots$ and f, respectively. Denote their distribution functions by $F_1, F_2, \dots$ and F, respectively. Suppose that $f_n(x) \xrightarrow{n\to\infty} f(x)$ for every $x \in \mathbb{R}$. Then $F_n(x) \xrightarrow{n\to\infty} F(x)$ for every $x \in \mathbb{R}$.*

(Hint: Use the Fatou lemma to show that $F(x) \le \liminf_{n\to\infty} F_n(x)$. For the other direction, consider $G(x) := 1 - F(x)$.)

7 Linear spaces, norm, inner product

7.1 Which of the spaces V below are linear spaces and why?

a.) $V := \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 + 2x_2 = 0\}$, with the usual addition and the usual multiplication by a scalar.

b.) $V := \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 + 2x_2 = 3\}$, with the usual addition and the usual multiplication by a scalar.

c.) $V := \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 \geq 0\}$, with the usual addition and the usual multiplication by a scalar.

d.) $V := \{f : (0, 1) \to \mathbb{R} \,|\, f \text{ is continuous and } |f| \leq 100\}$, with the usual addition and the usual multiplication by a scalar.

e.) $V := \{f : (0, 1) \to \mathbb{R} \,|\, f \text{ is continuous and bounded}\}$, with the usual addition and the usual multiplication by a scalar.

7.2 On the linear spaces V and W below, which of the given transformations $T : V \to W$ are linear and why?

a.) $V = \mathbb{R}^3$, $W = \mathbb{R}^2$, $T((x_1, x_2, x_3)) := (x_1, x_2 + x_3)$.

b.) $V = \mathbb{R}^3$, $W = \mathbb{R}^2$, $T((x_1, x_2, x_3)) := (x_1, 1 + x_3)$.

c.) $V = \mathbb{R}^3$, $W = \mathbb{R}^2$, $T((x_1, x_2, x_3)) := (x_1, x_2x_3)$.

d.) $V := \{f : (-1, 1) \to \mathbb{R} \,|\, f \text{ differentiable}\}$, with the usual addition and the usual multiplication by a scalar; $W := \mathbb{R}$; $T(f) := f'(0)$.

7.3 On the linear spaces V below, which of the given two-variable functions $B : V \to \mathbb{R}$ are bilinear forms? Which ones are symmetric and positive definite? Why?

a.) $V = \mathbb{R}^3$, $B((x_1, x_2, x_3), (y_1, y_2, y_3)) := x_1y_2 + x_2y_3 + x_3y_1$

b.) $V = \mathbb{R}^2$, $B((x_1, x_2), (y_1, y_2)) := x_1x_2 + y_1y_2$

c.) $V = \mathbb{R}^2$, $B((x_1, x_2), (y_1, y_2)) := x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2$

d.) $V := \{f : [-1, 1] \to \mathbb{R} \,|\, f \text{ is differentiable}\}$, with the usual addition and the usual multiplication by a scalar; $B(f, g) := \int_{-1}^{1} x^2 f(x)g(x)\,\mathrm{d}x$

e.) $V := \{f : [-1, 1] \to \mathbb{R} \,|\, f \text{ is differentiable}\}$, with the usual addition and the usual multiplication by a scalar; $B(f, g) := \int_{-1}^{1} x f(x)g(x)\,\mathrm{d}x$

f.) $V := \{f : [-1, 1] \to \mathbb{R} \,|\, f \text{ is differentiable}\}$, with the usual addition and the usual multiplication by a scalar; $B(f, g) := \int_{-1}^{1} f'(x)g(x)\,\mathrm{d}x$

7.4 Let V be an inner product space. Show that the function $N : V \to \mathbb{R}$ defined as $N(x) := \sqrt{\langle x, x\rangle}$ is indeed a norm (usually denoted as $||x|| = N(x)$).

8 Riesz representation theorem

8.1 Let V be an inner product space, and let d denote the natural metric on it (defined as $d(x, y) := ||x - y||$). Let $x \in V$, let $D \subset V$ be convex, and assume that $d(x, D) = R > 0$ (where $d(x, D) := \inf\{d(x, y) \,|\, y \in D\}$ is the distance of x and D). Find a number $C \in \mathbb{R}$ (possibly depending on R) such that if $u, v \in D$, $d(x, u) \leq R + \varepsilon$ and $d(x, v) \leq R + \varepsilon$ with some $\varepsilon < R$, then $d(u, v) \leq C\sqrt{\varepsilon}$. *(Hint: estimate the length of the longest line segment that fits in the shell $\{y \in V \mid R \leq d(x, y) \leq R + \varepsilon\}$. A two-dimensional drawing will help.)*

8.2 Let V be an inner product space, and let d denote the natural metric (defined as $d(x,y) := ||x-y||$).

a.) Let $a, c, x \in V$ with $x \neq c$. Calculate the distance of a from the line $\{c + t(x-c) \mid t \in \mathbb{R}\}$ using $||a-c||$, $||x-c||$ and $\langle a-c, x-c\rangle$.

b.) Let $E \subset V$ be a linear subspace and let $a \in V$. Suppose that $c \in E$ is such that $d(a,x) \geq d(a,c)$ for every $x \in E$ – which means that c is the point in E which is closest to a. Prove that E is orthogonal to $a-c$, meaning that $\langle x, a-c\rangle = 0$ for every $x \in E$.

8.3 Let V be an inner product space over $\mathbb{R}$ and let $f : V \to \mathbb{R}$ be a linear form. Let $E := \{y \in V \mid f(y) = 0\}$ be the null-space of f. Suppose that $f(a) = 1$, $c \in E$ and $a-c$ is orthogonal to E, meaning $(a-c)y = 0$ for every $y \in E$. Now, for any $x \in V$, find the $\lambda \in \mathbb{R}$ for which $x_1 := x - \lambda(a-c) \in E$. Use this to get the relation between $f(x)$ and $(a-c)x$.

8.4 Represent the following functions $f : V \to \mathbb{R}$ as multiplication by a fixed vector, whenever this is possible due to the Riesz representation theorem.

a.) $V = \mathbb{R}^{10}$ with the usual inner product, $f((x_1, \ldots, x_{10})) := x_5$ (evaluation at 5)

b.) $V = \mathbb{R}^{10}$ with the usual inner product, $f((x_1, \ldots, x_{10})) := x_6 - x_5$ (discrete derivative at 5).

c.) $V = \mathbb{R}^{10}$ with the usual inner product, $f((x_1, \ldots, x_{10})) := x_6 - 2x_5 + x_4$ (discrete second derivative at 5).

d.) $V = l^2 := \{x : \mathbb{N} \to \mathbb{R} \mid \sum_{i=1}^{\infty} x^2(i) < \infty\}$, with the inner product $x \cdot y := \sum_{i=1}^{\infty} x(i)y(i)$; $f(x) := \sum_{i=1}^{100} x(i)$.

e.) $V = l^2 := \{x : \mathbb{N} \to \mathbb{R} \mid \sum_{i=1}^{\infty} x^2(i) < \infty\}$, with the inner product $x \cdot y := \sum_{i=1}^{\infty} x(i)y(i)$; $f(x) := \sum_{i=1}^{\infty} x(i)$.

f.) $V = l^2 := \{x : \mathbb{N} \to \mathbb{R} \mid \sum_{i=1}^{\infty} x^2(i) < \infty\}$, with the inner product $x \cdot y := \sum_{i=1}^{\infty} x(i)y(i)$; $f(x) := \sum_{i=1}^{\infty} x^2(i)$.

g.) $V = L^2([0,1]) := \{x : [0,1] \to \mathbb{R} \mid \int_0^1 x^2(t)\,\mathrm{d}t < \infty\}$, with the inner product $x \cdot y := \int_0^1 x(t)y(t)\,\mathrm{d}t$; $f(x) := x(\frac{1}{2})$ (evaluation at $\frac{1}{2}$).

h.) $V = L^2([0,1]) := \{x : [0,1] \to \mathbb{R} \mid \int_0^1 x^2(t)\,\mathrm{d}t < \infty\}$, with the inner product $x \cdot y := \int_0^1 x(t)y(t)\,\mathrm{d}t$; $f(x) := x'(\frac{1}{2})$ (derivative at $\frac{1}{2}$).

i.) $V = L^2([0,1]) := \{x : [0,1] \to \mathbb{R} \mid \int_0^1 x^2(t)\,\mathrm{d}t < \infty\}$, with the inner product $x \cdot y := \int_0^1 x(t)y(t)\,\mathrm{d}t$; $f(x) := \int_{0.2}^{0.7} x(t)\,\mathrm{d}t$.

j.) $V = \{x : [0,1] \to \mathbb{R} \mid \int_0^1 x^2(t)\,\mathrm{d}t < \infty,\ f \text{ is differentiable}\}$, with the inner product $x \cdot y := \int_0^1 x(t)y(t)\,\mathrm{d}t$; $f(x) := x'(\frac{1}{2})$.

k.) $V = \{x : [0,1] \to \mathbb{R} \mid \int_0^1 x^2(t)\,\mathrm{d}t < \infty,\ f \text{ is continuous}\}$, with the inner product $x \cdot y := \int_0^1 x(t)y(t)\,\mathrm{d}t$; $f(x) := x(\frac{1}{2})$.

l.) $V = \{x : [0,1] \to \mathbb{R} \mid \int_0^1 x^2(t)\,\mathrm{d}t < \infty,\ f \text{ is continuous}\}$, with the inner product $x \cdot y := \int_0^1 x(t)y(t)\,\mathrm{d}t$; $f(x) := \int_{0.2}^{0.7} x(t)\,\mathrm{d}t$.

9 Radon-Nikodym theorem

9.1 Let $(X, \mathcal{F})$ be a measurable space and let μ, ν be σ-finite measures on it. Show that there is a countable partition $X = \dot{\bigcup}_i A_i$ such that $\mu(A_i) < \infty$ and $\nu(A_i) < \infty$ for every i. Use this to show that the special case of the Radon-Nikodym theorem for finite measures implies the general theorem (for σ-finite measures).

9.2 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X : \Omega \to \mathbb{R}^+$ be integrable and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Define $\nu : \mathcal{G} \to \mathbb{R}^+$ by $\nu(A) := \int_A X \, d\mathbb{P}$ (whenever $A \in \mathcal{G}$). Check that ν is a measure on $(\Omega, \mathcal{G})$. Show that Lebesgue measure on $\mathbb{R}$ is absolutely continuous w.r.t. counting measure (on $\mathbb{R}$), but it does not have a density. Why doesn't this contradict the Radon-Nikodym theorem?

10 Conditional expectation

10.1 Let X be a nonempty set and let $\mathcal{F}_i \subset 2^X$ be a σ-algebra for every $i \in I$, where I is some index set. I may be arbitrary (possibly much bigger that countable), but we assume $I \neq \emptyset$. Show that $\mathcal{F} := \bigcap_{i\in I} \mathcal{F}_i$ is also a σ-algebra. (Note that the assumption $I \neq \emptyset$ is important.)

10.2 Let $(\Omega, \mathcal{F})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be (Borel-)measurable. Let $(\mathcal{G}_i)_{i\in I}$ be the family of all σ-algebras over Ω such that X is $\mathcal{G}_i$-measurable, and let $\mathcal{G} := \bigcap_{i\in I} \mathcal{G}_i$. Show that $\mathcal{G}$ is the *smallest* σ-algebra for which X is measurable. (In what sense exactly is it the smallest?)

10.3 Let $(\Omega, \mathcal{F})$ be a probability space, let $X : \Omega \to \mathbb{R}$ be $(\mathcal{F}, \mathcal{B})$-measurable, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. Let $\sigma(X)$ be the smallest σ-algebra on Ω for which X is measurable. (This exists by the previous exercise.) This is called the *σ-algebra generated by X*. Show that

$$\sigma(X) = \{X^{-1}(B) \,|\, B \in \mathcal{B}\}.$$

10.4 Let $(\Omega, \mathcal{F})$ be a probability space, and let $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F}$ be sub-σ-algebras. We say that $\mathcal{F}_1$ and $\mathcal{F}_1$ are independent if any $A \in \mathcal{G}_1$ and $B \in \mathcal{G}_2$ are independent. Show that if the random variables X and Y are independent, then $\sigma(X)$ and $\sigma(Y)$ are independent.

10.5 Let $(\Omega, \mathcal{F})$ be a probability space, and let $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F}$ be sub-σ-algebras. Let X and Y be random variables, $X \in \mathcal{G}_1$, $Y \in \mathcal{G}_2$. Show that if $\sigma(X)$ and $\sigma(Y)$ are independent, then X and Y are independent.

10.6 Show that if X is a random variable, $f : \mathbb{R} \to \mathbb{R}$ measurable and $Y = f(X)$, then $\sigma(Y) \subset \sigma(X)$. Show an example when equality holds, and an example when not.

10.7 Show that if X, Y are independent random variables and $f, g : \mathbb{R} \to \mathbb{R}$ are measurable, then $f(X)$ and $g(Y)$ are also independent.

10.8 Show that the random variables $X, Y : \Omega \to \mathbb{R}$ are independent if and only if the (joint) distribution of the pair (X, Y) (which is a probability measure on $\mathbb{R}^2$) is the product of the distributions of X and Y.

10.9 Show that if X and Y are independent and integrable, then $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$.

10.10 Show that if the random variable X is independent of the σ-algebra $\mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}X$.

10.11 Let $\Omega = \{a, b, c\}$ and $\mathbb{P}$ the uniform measure on it. Let $X = \mathbf{1}_{\{c\}}$ and let $\mathcal{G} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$. Calculate $\mathbb{E}(X|\mathcal{G})$.

10.12 We roll two fair dice and let X, Y be the numbers rolled. Calculate $\mathbb{E}(X|X + Y)$.

10.13 Let $\Omega = [0, 1]^2$ and let $\mathbb{P}$ be Lebesgue measure on Ω. Let $X, Y : \Omega \to \mathbb{R}$ be defined as $X(u, v) = u$ and $Y(u, v) = \sqrt{u + v}$. Calculate $\mathbb{E}(Y|X)$.

10.14 Let U and V be independent random variables, uniformly distributed on $[0, 1]$. Calculate $\mathbb{E}(\sqrt{U + V}|U)$.

10.15 Let U and V be independent random variables, uniformly distributed on $[0,1]$. Calculate $\mathbb{E}(U+V|U-V)$.

10.16 Let U and V be independent random variables, uniformly distributed on $[0,1]$. Calculate $\mathbb{E}(\sqrt{U+V}|U-V)$.

10.17 Let X and Y be independent standard Gaussian random variables. Let $U = X+Y$ and $V = 2X - Y$. Calculate $\mathbb{E}(V|U)$. *(Hint: if W is independent of U, then $\mathbb{E}(W|U) = \mathbb{E}W$. If you choose $\lambda \in \mathbb{R}$ cleverly, then $W := V - \lambda U$ will be independent of U. (Since U and W are jointly Gaussian, to show independence it's enough to check that $Cov(U,W) = 0$.) Then write $V = \lambda U + W$.)*