Math 587 – Assignments 3 and 4


Exercises to hand in – Assignment 3

1. Read the proof of Minkowski’s inequality from the course text. Then prove Minkowski’s inequality, giving a little more detail than is given in the course text. State and prove versions of Hölder’s and Minkowski’s inequalities in the case \( p = \infty \) (and so \( q = 1 \) for Hölder).

2. Questions 4.6, 4.9, 7.1, 7.2 from the course text.

3. **(Shannon’s theorem.)** Suppose that \( X_1, X_2, \ldots \) simple, independent random variables such that for \( 1 \leq i \leq r < \infty \), \( \mathbb{P}[X_1 = v_i] = p(v_i) \) (for fixed \( r \geq 1 \), some \( v_1, \ldots, v_r \) and some non-negative \( p(v_1), \ldots, p(v_r) \) with \( \sum_{i=1}^{r} p(v_i) = 1 \)).

   Let \( H_n = H_n(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i) \). Show that

   \[
   -\frac{1}{n} \log H_n \xrightarrow{a.s.} h := \sum_{i=1}^{r} -p(v_i) \log p(v_i). 
   \]

   (Adapted from Billingsley : In information theory, \( v_1, \ldots, v_r \) are interpreted as the letters of an alphabet, \( X_1, X_2, \ldots \) are the successive letters produced by some information source, and \( h \) is the entropy of the source. Shannon’s theorem implies that for all \( \varepsilon > 0 \), for large \( n \) there is probability exceeding \( 1 - \varepsilon \) that the probability of the observed \( n \)-character sequence, or message, is in the range \( e^{-hn\pm\varepsilon n} \).)

Supplemental reading

1. Read the proofs of Hölder’s inequality and the Cauchy–Schwartz inequality from the book.

2. If you’re planning to take Math 589, read all of Chapter 6 (you may have to do this anyway as I may return to this later in term).

Supplemental exercises (do not hand in)

1. Find four random variables taking values in \( \{-1, 1\} \), so that any three are mutually independent but all four are not.

2. Let \( \Omega = \{1, 2, 3, 4\} \), \( \mathcal{F} = 2^\Omega \), and \( \mathbb{P}(\{i\}) = 1/4 \) for each \( i = 1, 2, 3, 4 \). Find \( \mathcal{A}_1 \subset \mathcal{F} \), \( \mathcal{A}_2 \subset \mathcal{F} \), such that \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) are independent but \( \sigma(\mathcal{A}_1) \) and \( \sigma(\mathcal{A}_2) \) are not.

3. Under the assumptions of Jensen’s inequality, show that if \( \phi \) is strictly convex then \( \phi(\mathbb{E}[X]) = \mathbb{E}[\phi(X)] \) if and only if \( X = \mathbb{E}[X] \) almost surely.
4. In this question we write $dx$ for Lebesgue measure on $([0, 1], B([0, 1]))$. Let $f : [0, 1] \to \mathbb{R}$ be a Borel function that is bounded and non-negative. Show that if 

$$\int_{[0,1]} x^j f(x)dx = a^{j+1}$$

for each $j = 0, 1, 2$ then $f = 0$ almost everywhere.

5. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and fix functions $f_n \in L^1(\mu)$, $n \geq 1$, with

$$\sum_{n \geq 1} \int |f_n|d\mu < \infty.$$

Define $f : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ by $f(\omega) = \sum_{n \geq 1} f_n(\omega)$. Show that $f(\omega) \in \mathbb{R}$ for almost every $\omega \in \Omega$, that $f \in L^1(\mu)$, and that

$$\int f d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$ 

Carefully justify steps involving the notation “almost everywhere”.

6. Suppose that $f \in L^1(\mu)$. Show that for all $\varepsilon > 0$ there is $\delta > 0$ such that for all $E \in \mathcal{F}$, if $\mu(E) < \delta$ then $\int_E |f|d\mu < \varepsilon$.

7. Show that if $f, g \in m\mathcal{F}$, $f, g \geq 0$ and $fg \geq 1$ almost everywhere, then

$$\left(\int f d\mu\right) \cdot \left(\int g d\mu\right) \geq 1.$$

Assignment 4, exercises to hand in

This assignment is about maximal inequalities and large deviations.

Let $X_i$, $1 \leq i \leq n$ be mutually independent, with mean 0 and variance 1. For $1 \leq i \leq n$ write $S_i = \sum_{1 \leq j \leq i} X_j$. Fix $t > 0$.

(a) Write the event that $|S_n| \leq t$ as a disjoint union, by considering the first time $i \leq n$ that $|S_i| \geq 2t$ (one possibility is that there is no such time).

(b) Use independence to show that

$$\mathbb{P}\left[\bigcap_{j<i} |S_j| < 2t, |S_i| \geq 2t, |S_n| \leq t\right] \leq \mathbb{P}\left[\bigcap_{j<i} |S_j| < 2t, |S_i| \geq 2t\right] \mathbb{P}[|S_{n-i}| \geq t]$$

(c) Show that

$$\mathbb{P}\left[\max_{1 \leq i \leq n} |S_i| \geq 2t\right] \leq \max_{1 \leq i \leq n} \mathbb{P}[|S_{n-i}| \geq t]$$

(d) Conclude that $\mathbb{P}\left[\max_{1 \leq i \leq n} |S_i| \geq 2t\right] \leq n/t^2$. Note that this bound is a factor of four weaker than Chebyshev’s inequality for sums, but gives a bound for $\max_{1 \leq i \leq n} |S_i|$ rather than just $|S_n|$. In the next part we get rid of the extra factor of four.
(e) For \(1 \leq i \leq n\) let \(B_i = \{|S_i| \geq t, |S_j| < t, 1 \leq j < i\}\). Using that \(S_n^2 1_{B_i} \geq (S_i^2 + 2S_i(S_n - S_i))1_{B_i}\), plus independence show that
\[
\mathbb{E}[S_n^2] \geq \sum_{1 \leq i \leq n} \mathbb{E}[S_i^2 1_{B_i}].
\]

(f) Use the inequality from (e) to show that
\[
\mathbb{P}
\left[
\max_{1 \leq i \leq n} |S_i| \geq t
\right]
\leq \frac{n}{t^2}
\]

(g) Generalize to the case where the \(X_i\) have mean \(\mu\) and variance \(\sigma^2\). (You do not have to write out proofs in detail, just reduce to the previously considered case.)

(h) Show that for all \(\lambda > 0\) and \(n \geq 1\), \(\mathbb{E}[e^{\lambda S_n}] \geq 1\).

(i) Show that if also \(|X_i| \leq K\) for all \(i\), then for any \(\lambda > 0\),
\[
\mathbb{P}
\left[
\max_{1 \leq i \leq n} S_i \geq t
\right]
\leq (\mathbb{E}[e^{\lambda(K-t)}])^n.
\]

(Hint.) Use a slight variant of the events \(B_i\) from (e), and bound \(\mathbb{E}[e^{\lambda S_n}]\) from below as in (e), starting from the equality
\[
\mathbb{E}[e^{\lambda S_n}] = \sum_{1 \leq i \leq n} \mathbb{E}[e^{\lambda S_n 1_{B_i}}]
\]
Use that \(e^{\lambda S_n} = e^{\lambda S_i e^{\lambda(S_n - S_i)}}\), plus independence and the result from (i).

(j) Suppose that the \(X_i\) are such that \(\mathbb{P}[X_i = 1] = 1/2 = \mathbb{P}[X_i = -1]\). Fix \(c \in (0, 1)\). Optimize the first bound from (i) over \(\lambda\) to give an explicit bound for
\[
\mathbb{P}[|S_n| \geq cn].
\]

(k) For \(X_i\) as in (j), use Stirling’s formula \(n! \sim \sqrt{2\pi n}(n/e)^n\) to find an approximate expression for \(\mathbb{P}[|S_n| = \lfloor cn \rfloor]\), and compare this with your answer from (j).

(l) (Hard, optional.) See if you can find a more general class of random variables for which optimizing the upper bound in (i) gives a bound that is correct to within \(e^{o(n)}\). In other words, if \(\lambda^*\) is the value of \(\lambda\) which optimizes the upper bound in (i), then for any \(\varepsilon > 0\) and all \(n\) sufficiently large, we should have
\[
(\mathbb{E}[e^{\lambda^*(X-t)}])^n \cdot e^{-\varepsilon n} \leq \mathbb{P}[S_n \geq t] \leq \mathbb{P}
\left[
\max_{1 \leq i \leq n} S_i \geq t
\right] \leq (\mathbb{E}[e^{\lambda^*(X-t)}])^n.
\]