HW1

Problem 1.1. The answer is $2^{(2^n)}$, which is not $4^n$ as somebody believed.

Problem 1.3. The event in question is
\[\{(0, 0, \omega_3, ..., \omega_n) : \omega_i \in \{0, 1\}\}\]
which contains $2^{n-2}$ points. By definition its probability is $2^{n-2}/2^n = 1/4$.

Pay attention to this “by definition”, not by intuition, by “independence”, or anything “natural”. You have to rely only on the definitions and facts established in this course.

Problem 1.4. Denote by $A$ the following sequence of length $j$
\[(0, ..., 0, 1),\]
Then the event in question is
\[\{(\omega_1, \omega_2, ...) : (\omega_1, ..., \omega_j) \in A\}\]
By definition its probability is $#A/2^j = 2^{-j}$.

Problem 1.5. $\Omega = \emptyset \in \mathcal{F}$.
\[\bigcup_{n=1}^{\infty} A_n \subseteq \bigcap_{n=1}^{\infty} A_n^c \in \mathcal{F}\]
if $A_n \in \mathcal{F}$. Hence
\[\bigcup_{n=1}^{\infty} A_n = (\bigcup_{n=1}^{\infty} A_n^c)^c \in \mathcal{F}.\]

Finally,
\[A \setminus B = A \cap B^c = A \cap B^c \cap \Omega \cap \ldots \in \mathcal{F}\]
as long as $A, B \in \mathcal{F}$.

Problem A. Introduce a function $f$ on $\{1, 2, ..., n\}$ by setting $f(k) = 1$ if $k$ is even and $f(k) = 0$ if $k$ is odd. Our event is the disjoint union of events that the number of heads is $0, 2, 4, ...$. Hence its probability is the sum of the corresponding probabilities and since we know that the probability to have exactly $m$ heads is $\binom{n}{m}2^{-n}$ the probability in question is
\[\sum_{k=1}^{n} \binom{n}{k}2^{-n}f(k)\]
Observe that $f(k) = (1/2)(1 + (-1)^k)$ and by the binomial formula
\[\sum_{k=1}^{n} \binom{n}{k} = 2^n, \ \sum_{k=1}^{n} \binom{n}{k}(-1)^k = (1 + (-1))^n = 0.\]
Taking this into account we get that our probability is $1/2$.

Problem B. The event in question is represented as the disjoint union of $A_{2n}$ that the first head occurred on flip number $2n$, $n = 1, 2, ...$. 
The event $A_{2n}$ is represented as one sequence of length $2n$ of zero’s and one’s, hence by definition

$$P(A) = \frac{\#A_{2n}}{2^{2n}} = 4^{-n}.$$ 

We conclude

$$P(A) = \sum_{n=1}^{\infty} 4^{-n} = 1/3.$$ 

Problem C. Since the intervals $(r, \infty)$ are open, $E \subset \mathcal{B}(\mathbb{R})$ and $\sigma(E) \subset \mathcal{B}(\mathbb{R})$.

To prove the opposite inclusion first note the following.

**Lemma 1.** Every nonempty open set in $\mathbb{R}^d$ is the countable union of open balls with rational centers and radii.

**Proof.** Let $U$ be an open set in $\mathbb{R}^d$. Since the case that $U = \mathbb{R}^d$ is trivial, assume that $U \neq \mathbb{R}^d$. The set of rational points in $\mathbb{R}^d$ is everywhere dense and countable. Let $\{x_1, x_2, \ldots\}$ be its subset consisting of all points which are in $U$. Take any rational $r_1, r_2, \ldots \in (0, \infty)$ such that $B_{r_n}(x_n) \subset U$ but $B_{2r_n}(x_n) \not\subset U$, where $B_r(x) = \{y \in \mathbb{R}^d : |x - y| < r\}$. Then

$$U = \bigcup_{n=1}^{\infty} B_{r_n}(x_n).$$

Indeed, if $x \in U$, then $B_r(x) \subset U$ for an $r > 0$ by definition. Now find $n$ such that $|x - x_n| < r/3$. Then $r_n \geq r/3$ since $B_{2(r/3)}(x_n) \subset U$. In this case $x \in B_{r_n}(x_n)$, which certainly proves the lemma.

By this lemma $\mathcal{B}(\mathbb{R})$ is generated by open intervals with rational endpoints. Take such an interval $(a, b)$ and observe that

$$(a, b) = (a, \infty) \cap (-\infty, b) = (a, \infty) \cap [b, \infty)^c.$$ 

This set is in $\sigma(E)$ because

$$[b, \infty) = \bigcap_{n=1}^{\infty} (b - 1/n, \infty)$$

and $\sigma(E)$ is closed under complementations and countable intersections.

Hence all open intervals with rational endpoints are elements of $\sigma(E)$ and the $\sigma$-fields they generate is a subset of $\sigma(E)$.

**HW2**

7.2. A). We have $\Omega \in \mathcal{A}$, so $\emptyset = \Omega \setminus \Omega \in \mathcal{A}$ and if $A \in \mathcal{A}$, then $A^c = \Omega \setminus A \in \mathcal{A}$, since $\mathcal{A}$ is a $\lambda$-system. It only remains to prove that if $A_1, A_2, \ldots \in \mathcal{A}$, then $\bigcup_n A_n \in \mathcal{A}$. Observe that

$$A_1 \cup A_2 = A_1 \cup (A_2 A_1^c)$$
and $A_2 A_i^c \in \mathcal{A}$ since $\mathcal{A}$ is a \( \pi \)-system and $A_1 \cup (A_2 A_i^c) \in \mathcal{A}$ since $\mathcal{A}$ is a \( \lambda \)-system. Hence, the union of two and therefore of any finitely many elements of $\mathcal{A}$ is in $\mathcal{A}$.

By observing that
\[ \bigcup_{n} A_n = \bigcup_{n=1}^{\infty} B_n, \]
where $B_n = \bigcup_{i \leq n} A_i$ and that $B_n \subset B_{n+1}$ we finish the proof.

7.8. Use the notation from above. Then
\[ \bigcup_{n} A_n = \bigcup_{n=1}^{\infty} (B_n \setminus B_{n-1}) \quad (B_0 := \emptyset), \]
and $B_n \setminus B_{n-1}$ are disjoint and are in $\mathcal{E}$. By \( \sigma \)-additivity and the assumption that $\bigcup_{n} A_n \in \mathcal{E}$ we have
\[ R\left( \bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} [P(B_n) - P(B_{n-1})] = \lim_{n \to \infty} P(B_n). \]

From lectures we know that if $A, B \in \mathcal{E}$ and $R(A) < \infty$, then
\[ R(A \cup B) = R(A) + R(B) - R(AB). \]
In that case $R(A \cup B) \leq R(A) + R(B)$. This inequality is also true obviously if $R(A) = \infty$. By adding terms one by one we see that
\[ P(B_n) \leq \sum_{i=1}^{n} R(A_n), \]
and we are done.

7.14. Let $\Sigma$ be the collection of regular sets. We are going to prove that
(i) $\Sigma$ is a \( \sigma \)-field, and
(ii) $B_r(x) \in \Sigma$.
Then by a property (proved before) of $\mathfrak{B}(\mathbb{R}^d)$ we have $\mathfrak{B}(\mathbb{R}^d) \subset \Sigma$, and this is exactly what we need.
Statement (ii) is almost trivial since, for every $n \geq 1,$
\[ \Gamma := B_r(x) \supset B_r(x) \supset \{x : \rho(x,y) \leq r - 1/n\} =: G_n, \]
where $\Gamma$ is open, the $G_n$ are closed and $\mu(\Gamma \setminus G_n) \to 0$ since the sets $\Gamma \setminus G_n$ are nested and their intersection is empty.
To prove (i), first notice that $\mathbb{R}^d \in \Sigma$ as a set open and closed simultaneously. Furthermore, the complement of an open (closed) set is a closed (respectively, open) set and if $G \supset B \supset \Gamma$, then $\Gamma^c \supset B^c \supset G^c$ with
\[ \Gamma^c \setminus G^c = G \setminus \Gamma. \]
This shows that if $B \in \Sigma$, then $B^c \in \Sigma$. It only remains to check that countable unions of elements of $\Sigma$ belong to $\Sigma$. 
Let $B_n \in \Sigma$, $n = 1, 2, 3, \ldots$, $\varepsilon > 0$, and let $G_n$ be open and $\Gamma_n$ be closed and such that

$$G_n \supset B_n \supset \Gamma_n, \quad \mu(G_n \setminus \Gamma_n) \leq \varepsilon 2^{-n}.$$ 

Define

$$B = \bigcup_n B_n, \quad G = \bigcup_n G_n, \quad D_n = \bigcup_{i=1}^n \Gamma_i.$$ 

Then $G$ is open, $D_n$ is closed, and obviously $G \setminus D_n$ are nested, so that

$$\lim_{n \to \infty} \mu(G \setminus D_n) = \mu(G \setminus D_\infty) \leq \sum_n \mu(G_n \setminus \Gamma_n) \leq \varepsilon.$$ 

Hence, for appropriate $n$ we have $\mu(G \setminus D_n) \leq 2 \varepsilon$, and this brings the proof to an end.

B) Let $\Omega^\mu_n$ and $\Omega^\nu_n$ be increasing sequences of elements of $\mathcal{E}$ tending to $\Omega$ as $n \to \infty$ and such that $\mu(\Omega^\mu_n) < \infty$ and $\nu(\Omega^\nu_n) < \infty$ for any $n$. Set $\Omega_n = \Omega^\mu_n \cap \Omega^\nu_n$. Then

$$\Omega_n \subset \Omega_{n+1}, \quad \Omega = \bigcup_n \Omega_n, \quad \mu(\Omega_n) = \nu(\Omega_n) < \infty.$$ 

Next, for each $n$, the collection of elements $A$ of $\mathcal{F}$ such that $\mu(\Omega^\mu_n A) = \nu(\Omega^\nu_n A)$ is a $\lambda$-system by almost obvious reasons, but you have to give these reasons. It contains $\mathcal{E}$, hence it contains $\mathcal{F}$ and $\mu(\Omega_n A) = \nu(\Omega_n A)$ for any $A \in \mathcal{F}$. By letting $n \to \infty$ and adding a simple argument based on the formula

$$\mu(A) = \sum_{n=1}^\infty \mu(A(\Omega_n \setminus \Omega_{n-1}))$$ 

and the fact that $A(\Omega_n \setminus \Omega_{n-1}) = \Omega_n[A(\Omega_n \setminus \Omega_{n-1})]$ one gets $\mu = \nu$ on $\mathcal{F}$.

C) Obviously any subset of a null set is a null set. Hence for any $X \subset \Omega$, $f^*(XA) + f^*(XA^c) = f^*(XA^c) \leq f^*(X)$. On the other hand $f^*(XA) + f^*(XA^c) \geq f^*(X)$ by subadditivity of $f^*$.

D) If $f^*(A) = \infty$, then $f^*(\Omega) = \infty$, and one can take $B = \Omega$.

If $f^*(A) < \infty$, then find $B_{nk} \in \mathcal{B}$ such that

$$B_n = \bigcup_{k=1}^\infty B_{nk} \supset A,$$

$$f^*(A) \geq \sum_{k=1}^\infty f(B_{nk}) - 1/n \geq f(B_n) - 1/n,$$

where the inequality follows from $\sigma$-additivity of $f$ (owing to one of the previous problems). Then $B := \bigcap_n B_n \supset A, B \in \mathcal{B}, f^*(A) \geq f(B_n) - 1/n \geq f(B) - 1/n$, and $f^*(A) \geq f(B)$. On the other hand, $f^*(A) \leq f^*(B) = f(B)$.
2.3. Let $A = \{ \omega : X(\omega) = Y(\omega) \}$ and $B$ be a measurable set in the space of values of $X$ and $Y$. Then $P(A \cap (X \in B)) = P(A \cap (Y \in B))$, $P(A^C \cap (X \in B)) \leq P(A^C) = 0$, and $P(X \in B) = P(A \cap (X \in B)) + P(A \cap (X \in B)) = P(A \cap (Y \in B))$. Similarly $P(Y \in B) = P(A \cap (Y \in B))$.

2.6. For $G \in \mathcal{G}$ we have $\{ \omega : X(\omega) \in G \} \in \mathcal{F}$. If $H \in \mathcal{H}$, then $G := \{ y : Y(y) \in H \} \in \mathcal{G}$. Hence $\{ \omega : Y(X(\omega)) \in H \} = \{ \omega : X(\omega) \in G \} \in \mathcal{F}$ and we are done.

7.19. For $s, t \geq 0$ we have $[0, t+s) = [0, t] \cup [t, t+s)$ and therefore $f(t+s) = f(t) + \mu([t, t+s)) = f(t) + f(s)$. Furthermore, $f(t)$ is left-continuous which is proved in the same way as the same property for distribution functions. Now if $p, q$ are integers $> 0$ and $t \geq 0$, then $f(tp/q) = f(t/q) + f(t(p-1)/q) = ... = pf(t/q)$, which for $t = q/p$ yields $f(1/p) = f(1)/p = 1/p$ and then with $t = 1$ we get $f(p/q) = p/q$. By the left continuity of $f$ we conclude that $f(t) = t$ for all $t \geq 0$. Therefore, for $a < b$, $\mu[a, b) = \mu[0, b-a) = b - a$ and by uniqueness $\mu$ is Lebesgue measure.

A) For any sets $M, N$ we have $M = MN \cup MN^c$. Hence, 

$$(B \cup N)^c = B^c N^c = (B^c C^c N^c) \cup (B^c C N^c) = (B^c C^c) \cup (B C^c N),$$

where the last equality follows from the fact that $N \subset C$. In addition $B^c C^c \in \mathcal{F}$, $CB^c N \subset C$ and $\mu(C) = 0$. It follows that $A^c \in \mathcal{F}$ if $A \in \mathcal{F}$. The fact that $\bigcup_n A_n \in \mathcal{F}$ if $A_n \in \mathcal{F}$ follows from the fact that $\mu(\bigcup_n C_n) = 0$ if $\mu(C_n) = 0$ for any $n$. We see that $\mathcal{F}$ is a $\sigma$-field indeed.

If $A \in \mathcal{F}$ admits two representations $B_1 \cup N_1$, $i = 1, 2$, with $B_i \in \mathcal{F}$, $N_i \subset C_i \in \mathcal{F}$ and $\mu(C_i) = 0$, then $B_1 \subset B_2 \cup C_2$, $\mu(B_1) \leq \mu(B_2) + \mu(C_2) = \mu(B_2)$. By symmetry $\mu(B_1) \geq \mu(B_2)$, so that $\mu(B_1) = \mu(B_2)$ and the definition $\mu(A) = \mu(B_1)$ is unambiguous.

The $\sigma$-additivity of $\mu$ on $\mathcal{F}$ is almost obvious although you are required to make an argument.

B) You get a solution by repeating what is written in Example 7.3 almost word for word.

C) We have

$$\{ \omega : \omega_{k_j}(\omega) = a_j \quad \forall j = 1, ..., n \} = \bigcap_{j=1}^n \{ \omega : \omega_{k_j}(\omega) = a_j \}$$

which is in $\mathcal{F} = \mathcal{B}[0,1)$ since $\omega_i$’s are random variables. Also

$$\{ \omega : \omega_{k_j}(\omega) = a_j \quad \forall j = 1, ..., n \} = \{ \omega : (\omega_1, ..., \omega_{kn}) \in A \},$$

where

$$A = \{(b_1, ..., b_{kn}) : b_i \in \{0, 1\}, b_{k_j} = a_j \quad \forall j = 1, ..., n \}.$$ 

Since $\#A = 2^{kn-n}$, the probability in question is $2^{-n}$.
D) We have
\[ P((X,Y) \in (a,1] \times (c,1]) = P(X > a, Y > c) = \sum_{n,m=1}^{\infty} P(A_{n,m}), \]
where
\[ A_{n,m} = (\omega_2 = a_1(a), \ldots, \omega_{2(n-1)} = a_{n-1}(a), \omega_{2n} = a_n(a), \omega_1 = a_1(c), \ldots, \omega_{2(m-1)-1} = a_{m-1}(c), \omega_{2m-1} = a_m(c)). \]

If \( a_n(a) = 1 \) or \( a_n(c) = 1 \), the set \( A_{n,m} \) is empty. However if \( a_n(a) = a_m(c) = 0 \), then
\[ A_{n,m} = (\omega_2 = a_1(a), \ldots, \omega_{2(n-1)} = a_{n-1}(a), \omega_{2n} = 1, \omega_1 = a_1(c), \ldots, \omega_{2(m-1)-1} = a_{m-1}(c), \omega_{2m-1} = 1) \]
and according to C) the probability of \( A_{n,m} \) is \( 2^{-n-m} \). Hence
\[ P((X,Y) \in (a,1] \times (c,1]) = \sum_{n,m=1}^{\infty} 2^{-n-m}(1 - a_n(a))(1 - a_m(c)) \]
\[ = \sum_{n=1}^{\infty} 2^{-n}(1 - a_n(a)) \sum_{m=1}^{\infty} 2^{-m}(1 - a_m(c)) = (1 - a)(1 - c). \]

You derive what is claimed by expressing \( (a,b] \times (c,d] \) in terms of sets like \( (a,1] \times (c,1] \).

**HW4**

3.5 Was done in class.

4.19 Denote
\[ Y_m = \sum_{n=1}^{m} X_n, \quad Y = \sum_{n=1}^{\infty} X_n. \]

Then \( Y_m \uparrow Y \) by definition, by additivity and definition
\[ EY_m = \sum_{n=1}^{m} EX_n \to \sum_{n=1}^{\infty} EX_n \]
as \( m \to \infty \), and it only remains to notice that by the monotone convergence theorem \( EY_m \to EY \).

A) We have \( z^\xi = \sum_{r=0}^{n} z^k I_{\xi=k} \) and \( EI_{\xi=k} = P(\xi = r) = \binom{n}{r} p^r q^{n-r}. \)
Hence
\[ Ez^\xi = \sum_{r=0}^{n} (pz)^r q^{n-r} = (pz + q)^n. \]
Then
\[ E\xi(\xi - 1) \cdots (\xi - k) = \sum_{r=0}^{n} r(r-1) \cdots (r-k)p^r q^{n-r} \]
\[ d^{k+1} = \sum_{r=0}^{n} (pz)^r q^{n-r} \mid_{z=1} = n(n-1) \cdots (n-k)p^{k+1}, \]

\[ E_\xi = np, \quad E_\xi(\xi - 1) = n(n-1)p^2, \quad E_\xi^2 = E_\xi + E_\xi(\xi - 1) = npq + (np)^2. \]

B) Let \( \Omega = \{1, 2, \ldots\} \), \( F \) be the collection of subsets of \( \Omega \), and \( \mu(A) = \#A \) for \( A \in F \). Then \( \mu \) is a measure and, if \( a_k \) is a nonnegative function on \( k \in \Omega \), then

\[ a_k = \sum_{r=1}^{\infty} a_r I_{\{r\}}(k), \quad \int_{\Omega} a_k \mu(dk) = \sum_{r=1}^{\infty} a_r \mu(\{k : k = r\}) = \sum_{r=1}^{\infty} a_r. \]

You see that what is required is a particular case of the monotone convergence theorem.

C) If \( x \in \lim_{n \to \infty} A_n \), then there exists an \( n_0 \) such that

\[ x \in \bigcap_{n=n_0}^{\infty} A_n \implies x \in \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_n \implies \lim_{n \to \infty} A_n \subseteq \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_n. \]

On the other hand if \( x \in \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_n \), there exists an \( n_0 \) such that

\[ x \in \bigcap_{n=n_0}^{\infty} A_n \implies \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_n \subseteq \lim_{n \to \infty} A_n. \]

The relation between the measures you have to prove follows directly from Fatou’s lemma and the fact that for any \( A \in F \)

\[ \int_{\Omega} I_A \mu(dx) = \mu(A). \]

D) Denote

\[ \Sigma = \{B : B \subset X, \xi^{-1}(B) \in F\}. \]

Then \( X \in \Sigma \) because \( \xi^{-1}(X) = \Omega \in F \). If \( B \in \Sigma \), then \( \xi^{-1}(B^c) = (\xi^{-1}(B))^c \in F \) and so on.

E) For any Borel \( B \subset [0, \infty] \) we have

\[ \mu(f \in B) = \mu(f^+ \in B) = \mu(-f \in B) = \mu(f^- \in B). \]

Hence \( \mu(f^+ \in A) = \mu(f^- \in A) \) for any Borel \( A \subset [-\infty, \infty] \), and since the integrals of measurable functions are uniquely determined by their distributions,

\[ \int_{\Omega} f^+ \mu(dx) = \int_{\Omega} f^- \mu(dx), \]

and we are done.
A) By Scheffé $E|\xi_n|^r - |\xi|^r \to 0$. On the set, where $|\xi_n - \xi| \geq 3|\xi|$ we have $|\xi_n| \geq 2|\xi|$, $|\xi_n - \xi|^r \leq 2^r|\xi_n|^r + 2^r|\xi|^r \leq C(|\xi_n|^r - |\xi|^r)$, where the last inequality is equivalent to $|\xi_n| \geq 2|\xi|$ for $C = (1 + 2^r)(1 - 2^{-r})^{-1}$. Hence,

$$E|\xi_n - \xi|^r I_{|\xi_n - \xi| \geq 3|\xi|} \leq CE^r |\xi_n|^r \to 0.$$ 

It only remains to add that $|\xi_n - \xi|^r I_{|\xi_n - \xi| < 3|\xi|} \leq 3^r|\xi|$ and by the dominated convergence theorem

$$E|\xi_n - \xi|^r I_{|\xi_n - \xi| \geq 3|\xi|} \to 0.$$ 

B) If $f = I_B$, where $B$ is a Borel set, the result follows from Problem B) of Homework 3. In the general case it is deduced from this particular case by standard arguments which you are supposed to provide.

C) We know that for any $\delta > 0$ there exists an open set $B \supset A$ such that $\mu(B \setminus A) \leq \delta$. The set $B$ is the disjoint union of open intervals, say $I_n$. Hence,

$$\mu(A) \leq \sum_n \mu(AI_n) \leq \varepsilon \sum_n \mu(I_n) = \varepsilon \mu(B) \leq \varepsilon \mu(A) + \delta.$$ 

We are done since $\varepsilon < 1$ and the inequality between the extreme terms holds for any $\delta > 0$.

D) First let $X = I_A$ and $Y = I_B$. They are uncorrelated iff $P(AB) = P(A)P(B)$, which is equivalent to the independence as shown in class. In the general case observe that, if $X$ and $Y$ are uncorrelated, then so are $c_1X + d_1$ and $c_2Y + d_2$ for any constants $c_1, d_1$. Then find these constants in such a way that these linear combinations take only two values 0 or 1, conclude that they are independent, and then expressing $X$ and $Y$ back show that they are independent as well.

E) We are given that for any nonnegative Borel $g$

$$Eg(X) = \int_{\mathbb{R}} f(x)g(x) \, dx.$$ 

If $g(x) = h(aX + b)$, then

$$Eh(aX + b) = \int_{\mathbb{R}} h(ax + b)g(x) \, dx = \frac{1}{|a|} \int_{\mathbb{R}} h(x)g((x - b)/a) \, dx,$$

where the last equality follows from Problem B). Hence the density of $aX + b$ is $(1/|a|)g((x - b)/a)$.

F) Take $0 < p \leq q < \infty$ and set $|X|^p = Y$ $r = q/p$ ($\geq 1$). Then by Jensen’s inequality, applied to $\phi(t) = |t|^r$, $(EY)^r \leq EY^r$ and this is exactly what is required.

G) See the solution of Problem B) of Homework 4.
9.36. The solution is obtained immediately after using the strong law of large numbers combined with the observation that 

\[ \lim_{{n \to \infty}} \frac{S_n}{M_n} = \lim_{{n \to \infty}} \frac{S_n/n}{M_n/n}. \]

9.45. We have \( P(X_n \leq 1) = 1/n \). Since the series of \( 1/n \) diverges, with probability one there is infinitely many \( n' \) such that \( X_{n'} \leq 1 \). Hence the probability in question is zero.

12.9. By the strong law of large numbers (a.s.) 

\[ \lim_{{n \to \infty}} \frac{1}{n} (X_1 + \ldots + X_n) = \mu = \lim_{{n \to \infty}} \frac{1}{n+1} (X_1 + \ldots + X_{n+1}) \]

where the last equality holds because \( n/(n+1) \to 1 \). By subtracting the extreme terms we get the result.

A) By Chebyshev for any constant \( c > 0 \)

\[ P(|\sum_{{n=1}}^m X_n - \sum_{{n=1}}^m \mu_n| \geq c) \leq \frac{1}{c^2} \text{Var} \sum_{{n=1}}^m X_n \leq \frac{1}{c^2} \sum_{{n=1}}^m \text{Var} X_n \leq \frac{M}{c^2} \sum_{{n=1}}^m \mu_n. \]

For

\[ c = \frac{1}{2} \sum_{{n=1}}^m \mu_n \]

we get that

\[ P(\sum_{{n=1}}^m X_n \leq (1/2) \sum_{{n=1}}^m \mu_n) \leq P(|\sum_{{n=1}}^m X_n - \sum_{{n=1}}^m \mu_n| \geq (1/2) \sum_{{n=1}}^m \mu_n) \]

\[ \leq 4M \left( \sum_{{n=1}}^m \mu_n \right)^{-1}. \]

It follows that

\[ P(\sum_{{n=1}}^{\infty} X_n < \infty) = \lim_{{m \to \infty}} P(\sum_{{n=1}}^{\infty} X_n \leq (1/2) \sum_{{n=1}}^m \mu_n) = 0, \]

which is what is required.

B) By Fubini and definition for \( r \in [a,b] \)

\[ \int_{{a}}^{r} \int_{{E}} f'(t,x) \mu(dx) dt = \int_{{E}} \left( \int_{{a}}^{r} f'(t,x) dt \right) \mu(dx) \]

\[ = \int_{{E}} [f(r,x) - f(a,x)] \mu(dx), \]
the result follows since the first expression is finite by assumption and

\[ \int_E |f(a, x)| \mu(dx) < \infty \]

by assumption.

C) First observe that \( F_n(x) \to F(x) \) at each point of continuity of \( F \) and thus at any point in \( \mathbb{R} \). Indeed, for any such point and \( \varepsilon > 0 \) there are \( x_1, x_2 \in \rho \) such that \( x_1 < x < x_2 \) and \( F(x_2) - F(x_1) \leq \varepsilon \). Then our assertion follows after examining

\[ F(x_1) = \lim_{n \to \infty} F_n(x_1) \leq \lim_{n \to \infty} F_n(x) \leq \lim_{n \to \infty} F_n(x_2) = F(x_2). \]

Also \( F_n(x-) \to F(x-) \) at any point in \( \mathbb{R} \), which is proved similarly.

Next, to prove the theorem it suffices to show that for any sequence \( x_n \in \mathbb{R} \) we have

\[ F_n(y_n) - F(y_n) \to 0. \]

If not all \( y_n = y \) starting from an \( n_0 \) (when we obviously obtain a contradiction), then either infinitely many of \( y_n \)'s are on the left of \( y \) or on the right of \( y \). Concentrate on the first possibility and then we have a sequence \( z_n \uparrow y \) such that \( F_n(z_n) - F(z_n) \to a \). However, this is impossible owing to the fact that for any \( \varepsilon > 0 \) for all large \( n \)

\[ F_n(y - \varepsilon) - F(z_n) \leq F_n(z_n) - F(z_n) \leq F_n(y) - F(z_n). \]

In the general case we may assume that the sequence \( F_n(x_n) - F(x_n) \) converges. If there is a subsequence of \( x_n \) which is bounded we are done by the above. In the remaining case there exists a subsequence \( y_n \) which tends either to \( \infty \) or to \( -\infty \). Consider the first case. Then for any \( m \) and large \( n \)

\[ F_n(m) - F(y_n) \leq F_n(y_n) - F(y) \leq 1 - F(y_n). \]

and the result follows.

D) \( P(X = r_n) \) is the jump of the distribution functo at \( r_n \). Since it is not zero, the distribution function is discontinuous at \( r_n \). Since the sum of jumps at rational points is one, there are no more jumps and the distribution function is continuous at irrational points.

**HW 7**

4.20. We are given that

\[ \sum_{n=1}^{\infty} |a_n| P(A_n) = \sum_{n=1}^{\infty} E(|a_n| I_{A_n}) = E \sum_{n=1}^{\infty} |a_n| I_{A_n} < \infty, \]

which implies that the series \( \sum a_n I_{A_n} \) converges (even absolutely) (a.s.). Fubini's theorem applied to the product of \( P \) and the counting measure on \( \{1, 2, \ldots\} \) allows us to interchange expectation and summation signs while computing \( EX \).
9.7. For \( t \geq 0 \) we have
\[
P(X \land Y > t) = P(X > t, Y > t) = P(X > t)P(Y > t) = e^{-2\lambda t},
\]
hence the distribution of \( X \land Y \) is exponential with parameter \( 2\lambda \).

14.1. The “only if” part is only worth commenting on. Observe that if \( F_n(r_i) \to F(r_i) \), \( r_1 < r_2 \), and \( r \) is a point lying between \( r_1 \) and \( r_2 \), then
\[
F(r_2) = \lim_{n \to \infty} F_n(r_2) \geq \lim_{n \to \infty} F_n(r) \geq \lim_{n \to \infty} F_n(r) \geq F(r_1).
\]
If \( r \) is a point of continuity of \( F \), then by sending \( r_1 \) and \( r_2 \) along our dense subset to \( r \) we conclude
\[
F(r) \geq \lim_{n \to \infty} F_n(r) \geq \lim_{n \to \infty} F_n(r) \geq F(r),
\]
which is the desired result.

14.14. It suffices to observe that for any function \( f \) on \( X = \{0, 1, 2, \ldots, n\} \) we have
\[
\int_X f(x)Q_p(dx) = \sum_{k=0}^{n} f(k) \binom{n}{k} p^k(1-p)^k,
\]
which shows that the integral is a continuous function of \( p \) and after that use the portmanteau theorem.

14.42. We have a sequence of distribution functions \( F, F_n, n = 1, 2, \ldots \), such that every its subsequence has a further subsequence which converges to \( F \) at each point of continuity of \( F \). Let \( x_0 \) be a point of continuity of \( F \). Then the sequence of numbers \( F_n(x_0), n = 1, 2, \ldots \), is such that every its subsequence has a further subsequence which converges to \( F(x_0) \). From the properties of real line it follows then that \( F_n(x_0) \to F(x_0) \) as \( n \to \infty \). Since this holds at every point of continuity of \( F \), we are done.

A) By the strong law of large numbers (a.s.)
\[
\ln \left( \prod_{k=1}^{n} X_k \right)^{1/n} = \frac{1}{n} \sum_{k=1}^{n} \ln X_k \to \int_{0}^{1} \ln x \, dx = ... 
\]

B) (i) \( \implies \) (ii). Let \( x_n \in K, n = 1, 2, \ldots \), and take a 1-net \( x_1 \), \ldots, \( x_{k_1} \). Then in at least one closed ball \( B_1(x_i^1) \) there are infinitely many elements of the sequence \( x_n \) (maybe not all distinct). Let it be \( B_1(x_i^1) \). Then take a 1/2-net \( x_1^2, \ldots, x_{k_2}^2 \), and observe that in at least one intersection \( B_1(x_i^1) \cap B_{1/2}(x_i^2) \), there are infinitely many elements of the sequence \( x_n \). Let it be \( B_1(x_i^1) \cap B_{1/2}(x_i^2) \). Then take a 1/4-net and so on.

In this way we construct a sequence \( y_1 := x_{i_1}^1, y_2 := x_{i_2}^2, \ldots \) such that \( \rho(y_m, y_{m+1}) \leq 3/2^m \) and in the \( 2^{-m} \)-neighborhood of \( y_m \) there is at least one element of the sequence \( x_n \). By the triangle inequality for \( k < m \)
\[
\rho(y_k, y_m) \leq \rho(y_k, y_{k+1}) + \ldots + \rho(y_{m-1}, y_m) \leq 3/2^k,
\]
which tends to zero as \( k, m \to \infty \). Since \( X \) is complete and \( K \) is closed, there is a point \( y_0 \in K \) such that \( y_m \to y_0 \). Furthermore obviously, in any neighborhood of \( y_0 \), there is at least one element of the sequence \( x_n \). Now to extract a convergent subsequence it suffices to take points \( x_{n_k} \), which are in \( 1/k \)-neighborhoods of \( y_0 \).

(ii) \( \implies \) (i). Assume the contrary: there is an \( \varepsilon > 0 \) such that no matter which points \( x_1, \ldots, x_n \) we take there always exist a point \( x_{n+1} \) which lies at the distance larger than \( \varepsilon \) from any of those points. Then pick \( x_1 \) arbitrarily, and choose \( x_2 \) so that \( \rho(x_1, x_2) \geq \varepsilon \) and \( x_3 \) so that \( \rho(x_1, x_3), \rho(x_2, x_3) \geq \varepsilon \) and so on. We will have a sequence \( x_n, n = 1, 2, \ldots \), such that \( \rho(x_n, x_{m}) \geq \varepsilon \) for any \( n, m \). Obviously it does not contain any convergent subsequence, which is the desired contradiction.

**Solutions for the final**

1. Denote \( Y = X(A \cup B) \). Then, since \( A \) is a \( \nu \)-set, we have

\[
\nu(Y) = \nu(YA) + \nu(YA^c).
\]

It only remains to note that, since \( AB = \emptyset \),

\[
YA = X(A \cup B)A =XA, \quad YA^c = X[(A \cup B) \setminus A] = XB.
\]

We got even more than required: the stated equality is true as long as (only) \( A \in \Sigma \).

2. The additivity is proved in the same way as in the case of right-continuous functions. If \( a \) is a point of discontinuity of \( F \), then

\[
\lim_{n \to \infty} R((a, a + 1/n)]) = F(a+) - F(a) > 0
\]

whereas the intersection of all \( (a, a + 1/n] \) is empty, which is incompatible with \( \sigma \)-additivity.

3. If \( \beta \) is the characteristic function of \( \xi \), then \( \beta(c_k t) \) is the characteristic function of \( c_k \xi \) and

\[
\sum_{k=1}^{n} a_k \beta(c_k t)
\]

is the characteristic function of the random variable \( c \xi \) where \( c \) is a random variable independent of \( \xi \) and taking values \( c_k \) with probabilities \( a_k \).

To prove Polya’s theorem, for each \( n = 1, 2, \ldots \), take points \( (k/n, \gamma(k/n)) \), \( k = 0, 1, \ldots, n^2 \), connect them with straight segments extending the graph of such obtained approximation \( \gamma_n \) of \( \gamma \) beyound \( n \) to be \( \gamma(n) \). Extend \( \gamma_n \) for \( t \leq 0 \) setting \( \gamma(t) = \gamma(-t) \).

Then observe that \( \gamma_n = \gamma^1_n + \eta_n \), where \( \eta(t) \) is an even function having the triangular shape equal to \( \gamma(n) \) for \( t \geq n \) and whose graph on \( [0, n] \) is a straight segment passing through \( ((n^2 - 1)/n, \gamma((n^2 - 1)/n)) \) and \( (n, \gamma(n)) \).

Since \( \gamma \) is convex on \( [0, \infty) \), we have that \( \gamma^1_n \geq 0 \) is convex and the number of breaks in its graph is smaller by one than that of the graph of \( \gamma_n \). Moreover by the above \( \eta_n \) a characteristic function. By using the induction on the
number of breaks, we get that $\gamma_n$ are characteristic functions. Their limit, $\gamma$ is continuous at the origin and thus is also the characteristic function of a probability distribution.

In case that $\gamma$ has two continuous derivatives on $(0, \infty)$, $\gamma'(0) = -\infty$, $\gamma(\infty) = 0$, and $\gamma'' > 0$ on $(0, \infty)$ one can give a formula for the distribution. It turns out that for $t \geq 0$

$$\gamma(t) = \int_0^\infty (\alpha'(x) - t) + dx,$$

where $\alpha$ is the Young's transform of $\gamma$:

$$\alpha(x) = \inf_{t \geq 0} [\gamma(t) + tx].$$

To prove this observe that for $x > 0$ the inf is obtained at a single point where $\gamma'(t) = -x$ ($\gamma'$ is a strictly increasing negative function). By the inverse function theorem the equation $\gamma'(t) = -x$ has a solution $t = t(x)$ which is a strictly decreasing differentiable function. Notice that

$$\alpha(x) = \gamma(t(x)) + t(x)x, \quad \alpha'(x) = t(x), \quad t(0) = \infty.$$  

Hence the right-hand side of (1) is the integral of $\alpha'(x) - t$ over $(0, x(t))$, where $x(t)$ is such that $\alpha'(x(t)) = t$, that is $t(x(t)) = t$ or $x(t) = -\gamma'(t)$. It follows that the right-hand side of (1) equals

$$\alpha(x(t)) + t\gamma'(t),$$

which is $\gamma(t)$ since due to the first equation in (2)

$$\alpha(x(t)) = \gamma(t) + t(-\gamma'(t)).$$

Then, of course, $\gamma$ is the characteristic function of $\xi/t(c)$, where $c$ is a random variable independent of $\xi$ and having density $\alpha'$.

4. You can use the dominated convergence theorem applied for counting measures as in the solution of Problem 4.20. Alternatively, observe that for any $\varepsilon > 0$ there exists a $k_0$ such that for any $n$

$$\sum_{k=k_0}^{\infty} |a_k^n| \leq \sum_{k=k_0}^{\infty} |b_k^n| \leq \varepsilon.$$ 

This, in particular, implies that any partial sum and the series

$$\sum_{k=k_0}^{\infty} |a_k| \leq \varepsilon.$$ 

It follow that

$$\lim_{n \to \infty} \sum_k a_k^n \leq \lim_{n \to \infty} \sum_{k=1}^{k_0} a_k^n + \varepsilon = \sum_{k=1}^{k_0} a_k + \varepsilon \leq \sum_{k=1}^{\infty} a_k + 2\varepsilon.$$
By replacing 3\(^n\) with \(-3^n\) we get
\[
\lim_{n \to \infty} \sum_{k} 3^n \geq \sum_{k=1}^{\infty} a_k - 2\varepsilon
\]
and hence the result.

5. Assume the contrary: there exist a \(\varepsilon > 0\) and a sequence of \(A_n \in \mathcal{F}\) are such that \(\mu(A_n) \to 0\), and
\[
\int_{\Omega} f(x) I_{A_n}(x) \mu(dx) \geq \varepsilon.
\]
Just by rarifying if needed the sequence \(A_n\), we may assume that \(\mu(A_n) \leq 2^{-n}\). Then set
\[
B_n = \bigcup_{m=n}^{\infty} A_m.
\]
Obviously, \(I_{B_n} \downarrow 0\). Hence by the dominated convergence theorem
\[
\int_{\Omega} f(x) I_{A_n}(x) \mu(dx) \leq \int_{\Omega} f(x) I_{B_n}(x) \mu(dx) \to 0,
\]
which yields the desired contradiction.

6. The solution follows from the formula
\[
P(X = a) = F(a) - F(a-),
\]
which shows a one-to-one correspondence between the points of discontinuity and concrete values of \(X\).

7. Introduce \(Y_n = X_n I_{X_n \leq 1}\) and observe that by Borel-Cantelly for almost any \(\omega\), \(X_n(\omega) = Y_n(\omega)\) for all large \(n\), so that if
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = c \quad (a.s.),
\]
then
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_k = c \quad (a.s.).
\]
Next, \(\text{Var} Y_n \leq EY_n^2 \leq EY_n : = \mu_n = (1 - 2^{-n})/2\) and
\[
\lim_{n \to \infty} \frac{\mu_1 + \ldots + \mu_n}{n} = 1/2,
\]
which by the strong law of large numbers (Lemma 17.10 of Lecture Notes) implies that the limit in question exists with probability one and \(c = 1/2\).

8. The hint provides the solution.

9. If we assume that it \(x(\cdot)\) is in the intersection of \(\Gamma_n\), then
\[
x(0) = \lim_{m \to \infty} x(2/(4\pi m + \pi))
\]
\[
= \lim_{m \to \infty} \left[ x(2/(4\pi m + \pi)) - f(2/(4\pi m + \pi)) \right] + 1 \geq 1/2,
\]
\[ x(0) = \lim_{m \to \infty} x(2/(4\pi m - \pi)) \]
\[ = \lim_{m \to \infty} [x(2/(4\pi m - \pi)) - f(2/(4\pi m - \pi))] - 1 \leq -1/2 \]
lead to a contradiction.

10. By Scheffé’s theorem we have
\[ \int_X |f_n - f| \mu(dx) \to 0 \]
and then for any Borel bounded (with no restrictions on the point of discontinuity) \( g \) we have
\[ |\int_X g Q_n(dx) - \int_X g Q(dx)| = |\int_X g(f_n - f) \mu(dx)| \]
\[ \leq \sup |g| \int_X |f_n - f| \mu(dx) \to 0. \]

11. For \( x \in C[0, 1] \) denote
\[ f(x) = \max_{[0,1]} x(t). \]
Since
\[ |\max_{[0,1]} x(t) - \max_{[0,1]} y(t)| \leq \max_{[0,1]} |x(t) - y(t)| = \rho(x, y), \]
the function \( f \) is continuous. Finally for any continuous bounded real-valued function \( g \) on \( \mathbb{R} \), the function \( g(f(x)) \) is a bounded continuous function on \( X \). By assumption
\[ Eg(f(X_n)) = \int_X g(f(x)) Q_{X_n}(dx) \to \int_X g(f(x)) Q_X(dx). \]
It only remains to observe that by the change of variables formula, for instance,
\[ Eg(f(X_n)) = \int_{\mathbb{R}} g(y) Q_{f(X_n)}(dy). \]

12. The characteristic function of
\[ \xi_k = \sum_{n=1}^{k} \frac{\varepsilon_n}{2^n} \]
is the product of the characteristic functions of \( \varepsilon_n 2^{-n} \), that is
\[ \beta_{\xi_k}(t) = \prod_{n=1}^{k} \cos \frac{t}{2^n}. \]
By using the hint
\[ \beta_{\xi_k}(t) \sin \frac{t}{2^k} = \frac{1}{2} \beta_{\xi_{k-1}}(t) \sin \frac{t}{2^{n-1}} = \cdots = \frac{1}{2^k} \sin t. \]
Hence for $t \neq 0$

$$\beta_\xi(t) = \lim_{k \to \infty} \beta_\xi_k(t) = \frac{\sin t}{t} \lim_{k \to \infty} \frac{t}{2\pi} \frac{\sin \frac{t}{2\pi}}{t} = \frac{\sin t}{t}.$$

13. We are given that for any $\varepsilon > 0$ there exists $\delta > 0$ such that $|u_n(x) - u_n(y)| \leq \varepsilon$ whenever $|x - y| \leq \delta$, $x, y \in [0, 1]$, and $n \geq 1$. By letting $n \to \infty$ we get that $|u(x) - u(y)| \leq \varepsilon$ whenever $|x - y| \leq \delta$, which means that $u$ is a continuous and even uniformly continuous function.

Now assume that what we want to prove is wrong. Then there exist an $\varepsilon > 0$ and a subsequence $n'$ and points $x_{n'} \in [0, 1]$ such that $|u_{n'}(x_{n'}) - u(x_{n'})| \geq 4\varepsilon$.

By the properties of real line there exists a subsequence $\{n(k)\}$ of $\{n'\}$ and a point $x_0 \in [0, 1]$ such that $x_{n(k)} \to x_0$ as $k \to \infty$. Then for $k$ large enough we have $|x_0 - x_{n(k)}| \leq \delta$, where $\delta$ is taken from above, and hence

$$4\varepsilon \leq \lim_{k \to \infty} |u_{n(k)}(x_{n(k)}) - u(x_{n(k)})| \leq \lim_{k \to \infty} |u_{n(k)}(x_{n(k)}) - u_{n(k)}(x_0)|$$

$$+ \lim_{k \to \infty} |u_{n(k)}(x_0) - u(x_0)| \leq 2\varepsilon,$$

which is the desired contradiction.

14. By preceding theorems we know that the sequence $Q_n$, $n \geq 1$, converges weakly to a probability distribution and that the sequence is tight. As in the case of the above Problem 13 to show the uniform convergence in question it suffices only to show that the sequence of characteristic functions is equicontinuous. This could be done either by using the continuity theorem which shows the equicontinuity on any ball, or using the following argument showing the equicontinuity on $\mathbb{R}^d$.

By tightness, for any $\varepsilon > 0$ there is an $R < \infty$ such that $Q_n(|x| \geq R) \leq \varepsilon$ for all $n$. Also recall that for real numbers $\alpha, \beta$

$$|e^{i\alpha} - e^{i\beta}| = |e^{i(\alpha - \beta)} - 1| = \left| \int_0^{\alpha - \beta} e^{it} dt \right| \leq |\alpha - \beta|.$$

Hence

$$\left| \int_{\mathbb{R}^d} e^{i(\lambda, x)} Q_n(dx) - \int_{\mathbb{R}^d} e^{i(\mu, x)} Q_n(dx) \right| \leq \int_{\mathbb{R}^d} |e^{i(\lambda, x)} - e^{i(\mu, x)}| Q_n(dx)$$

$$\leq 2\varepsilon + \int_{|x| \leq R} |\lambda - \mu| R Q_n(dx) \leq 2\varepsilon + |\lambda - \mu| R,$$

and the rest is trivial.