Solution for Note 2 of 275A Probability Theory

shuyu qin

August 2023

1 Selected Exercises in Note 2

Exercise 3

Proof. We use strong induction on $|A|$. For convenience, we identify the index set A of cardinality n with the set $\{1, 2, \ldots, n\}$. The base cases for $|A| = 1, 2$ are established already, now suppose the conclusions hold for $|A| = 1, 2, \ldots, n$ up to some natural number $n > 2$. That is, up to some natural number $n > 2$, there exists a unique probability measure $\prod_{i=1}^{n} \mu_i$ with the given property, and for any partition of the set $\{1, 2, \ldots, n\}$ into two sets, this probability measure equals the product measure determined by this partition.

Now consider when $|A| = n + 1$. For any set E of the form $\prod_{i=1}^{n+1} E_i$ with $E_i \in \mathcal{F}_i$, E can be written as $(\prod_{i=1}^{n} E_i) \times E_{n+1}$. Hence by our induction hypothesis and the base case there exists a unique probability measure $\prod_{i=1}^{n} \mu_i \times \mu_{n+1}$ on
$(\prod_{i=1}^{n} \Omega_i \times \Omega_{n+1}, \prod_{i=1}^{n} \mathcal{F}_i, \times \mathcal{F}_{n+1})$, with the property that

$(\prod_{i=1}^{n} \mu_i \times \mu_{n+1})(E) = (\prod_{i=1}^{n} \mu_i)(\prod_{i=1}^{n} E_i)\mu_{n+1}(E_{n+1}) = \prod_{i=1}^{n} \mu_i(E_i) \cdot \mu_{n+1}(E_{n+1})$,

by definition, this probability measure is exactly $\prod_{i=1}^{n+1} \mu_i$. Finally, for any partition of the set $\{1, 2, \ldots, n+1\}$ into two sets, we see again from what just shown and the induction hypothesis that this probability measure equals the product measure determined by this partition. $\square$

Exercise 4

Proof. By definition, the Stieltjes measure function of the product measure $\mu_1 \times \ldots \times \mu_n$ maps $(t_1, \ldots, t_n)$ to $\mu_1 \times \ldots \times \mu_n(\prod_{i=1}^{n}(-\infty, t_i]) = \prod_{i=1}^{n} \mu_i((-\infty, t_i]) = F(t_1) \ldots F(t_n)$. The uniqueness of the product measure and the Stieltjes measures associated to the F_i's then guarantee that this measure is unique. $\square$

Exercise 5

Proof. We can express $\mathbf{R}$ as a countable union of disjoint sets of measure 1 by writing $\mathbf{R} = \bigcup_{n\in\mathbf{Z}} I_n$, where $I_n := [n, n+1)$. Then for any Borel sets $E_1, E_2 \subset \mathbf{R}$, we have $E_1 = \bigcup_{n\in\mathbf{Z}} E_1 \cap I_n$, $E_2 = \bigcup_{m\in\mathbf{Z}} E_2 \cap I_m$, and $E_1 \times E_2 = \bigcup_{(n,m)\in\mathbf{Z}^2} F_{n,m}$, where $F_{n,m} := (E_1 \cap I_n) \times (E_2 \cap I_m)$.

By Theorem 1, there is a unique probability measure $m|_{I_n} \times m|_{I_m}$ on $(I_n \times I_m, \mathcal{B}[I_n] \times \mathcal{B}[I_m])$ such that $m|_{I_n} \times m|_{I_m}(F_{n,m}) = m(E_1 \cap I_n)m(E_2 \cap I_m)$. By characterization of the Lebesgue measure, $m|_{I_n} \times m|_{I_m}$ is exactly m^2(restricted to $I_n \times I_m$). By countable additivity, we thus have:

$m^2(E_1 \times E_2) = \sum_{(n,m)\in\mathbf{Z}^2} m^2(F_{n,m}) = \sum_{(n,m)\in\mathbf{Z}^2} m(E_1 \cap I_n)m(E_2 \cap I_m) = m(E_1)m(E_2)$,

where we have used Tonelli's theorem for series in the last step. □

Exercise 7

Proof. Suppose we have two measures μ_A and μ'_A on $\mathcal{F}_A$ that are product measures of $\{\mu_i : i \in A\}$ in the sense that

$\mu_A(\pi_B^{-1}(E_B)) = \mu'_A(\pi_B^{-1}(E_B)) = \mu_B(E_B)$

for all finite subsets B of A and all E_B in $\mathcal{F}_B$, where $\mu_B := \prod_{i\in B} \mu_i$. Let $\mathcal{F}$ be the collection of all $E \in \mathcal{F}_A$ such that $\mu_A(E) = \mu'_A(E)$, then $\mathcal{F}$ contains all sets of the form $\pi_B^{-1}(E_B)$ where $B \subset A$ is a finite subset. That is, $\mathcal{F}$ contains the collection $\mathcal{A}$ of elementary sets. But $\mathcal{A}$ is a Boolean algebra that generates $\mathcal{F}_A$, and from continuity from above and below wee see that $\mathcal{F}$ is a monotone class. By the monotone class lemma, we conclude that $\mathcal{F}$ is all of $\mathcal{F}_A$, hence $\mu_A = \mu'_A$. □

Exercise 9

Proof. Note that the compatibility condition is needed for the existence of μ_A. As $\pi_B^{-1}(\pi_{B\to C}^{-1}(E_C)) = \pi_C^{-1}(E_C)$, so if μ_A is a product measure on $(\Omega_A, \mathcal{F}_A)$ that extends measures on all finite product spaces in the sense of (7), then on the one hand we have $\mu_A(LHS) = \mu_B(\pi_{B\to C}^{-1}(E_C))$, on the other hand we have $\mu_A(RHS) = \mu_C(E_C)$.

Now suppose we have two measures μ_A and μ'_A satisfying the extension property (7), following the same argument as in Exercise 7 gives the uniqueness of μ_A. □

Note on Lemma 9:

1. Points in $[0,1]$ that has two binary expansions must be of the form $0.\sigma 000\ldots$ and $0.\sigma 111\ldots$, where σ is a finite sequence, and there are countably many such points.

2. If we manage to show that (X,d) is Borel isomorphic to a Borel subset of $[0,1]^{\mathbf{N}}$(and hence Borel isomorphic to a Borel subset A of [0, 1]), then for any Borel subset Ω of X, Ω is Borel isomorphic to a Borel subset $B \subset A \subset [0,1]$ by the same isomorphism. That is, Ω is standard Borel.

3. To see that the map $x \mapsto (\dfrac{d(x,q_i)}{1+d(x,q_i)})_{i\in\mathbf{N}}$ is injective, let x and y be two distinct points in X, since the sequence (q_n) is dense, there exists an open ball centered at x with arbitrarily small radius(smaller than $d(x,y)/2$, for instance) containing some q_i, then $d(x,q_i) < d(y,q_i)$. In particular, the point $(\dfrac{d(x,q_i)}{1+d(x,q_i)})_{i\in\mathbf{N}}$ is distinct from the point $(\dfrac{d(y,q_i)}{1+d(xy,q_i)})_{i\in\mathbf{N}}$.

4. The map $x \mapsto (\dfrac{d(x,q_i)}{1+d(x,q_i)})_{i\in\mathbf{N}}$ is measurable since the individual map $x \mapsto \dfrac{d(x,q_i)}{1+d(x,q_i)}$ for each $i \in \mathbf{N}$ is measurable(being the ratio of two continuous maps).

5. To see that the image of the map $x \mapsto (\dfrac{d(x,q_i)}{1+d(x,q_i)})_{i\in\mathbf{N}}$ is the closure of points $(\frac{d(q_j,q_i)}{1+d(q_j,q_i)})_{i\in\mathbf{N}}$ in the uniform norm(hence a Borel subset of $[0,1]$), note that for any $x \in X$, there exists some q_j arbitrarily close to x, thus making $|\dfrac{d(x,q_i)}{1+d(x,q_i)} - \dfrac{d(q_j,q_i)}{1+d(q_j,q_i)}|$ arbitrarily small for all $i \in \mathbf{N}$. That is, the point $(\frac{d(x,q_i)}{1+d(x,q_i)})_{i\in\mathbf{N}}$ is arbitrarily close to the point $(\frac{d(q_j,q_i)}{1+d(q_j,q_i)})_{i\in\mathbf{N}}$ in the uniform norm.

Exercise 13

Proof. We can use Theorem 10 as a "black box" to establish this Exercise. Let B be a box in $\mathbf{R}^n$, define an increasing sequence $E_1 \subset E_2 \subset \ldots$ of measurable sets in $\mathbf{R}^{n+2}$ by setting $E_m := B \times \mathbf{R} \times [-m,m]$ for each m. Then from continuity from below one has

$\mu_{n+2}(B\times\mathbf{R}^2) = \mu_{n+2}(\bigcup_{m=1}^{\infty} E_m) = \lim_{m\to\infty} \mu_{n+2}(E_m) = \lim_{m\to\infty} \mu_{n+1}(B\times[-m,m]) = \mu_{n+1}(B\times\mathbf{R}) = \mu_n(B)$.

By iteration, one obtains that for any $m \le n$, $\mu_n(B \times \mathbf{R}^{n-m}) = \mu_m(B)$ for all boxes $B \subset \mathbf{R}^m$(Including the empty box). Also, if $\mu_n(B \times \mathbf{R}^{n-m}) = \mu_m(B)$ for some box $B \subset \mathbf{R}^m$, as we are dealing with probability measures, we have $\mu_n(\overline{B} \times \mathbf{R}^{n-m}) = 1 - \mu_n(B \times \mathbf{R}^{n-m}) = 1 - \mu_m(B) = \mu_m(\overline{B})$. Finally, let $B_1, B_2, \ldots \subset \mathbf{R}^m$ be boxes, then by continuity from below we get: $\mu_n(\bigcup_{i=1}^{\infty} B_i \times \mathbf{R}^{n-m}) = \lim_{k\to\infty} \mu_n(\bigcup_{i=1}^{k} B_i \times \mathbf{R}^{n-m}) = \lim_{k'\to\infty} \mu_n(\bigcup_{i=1}^{k'} B_i^{'} \times \mathbf{R}^{n-m}) = \sum_{i=1}^{\infty} \mu_m(B_i^{'}) = \mu_m(\bigcup_{i=1}^{\infty} B_i)$, where we express the union of k boxes B_i as the union of k' disjoint boxes $B_i^{'}$. As the Borel σ-algebra on $\mathbf{R}^m$ is generated by the boxes, this implies the compatibility condition on μ_n that $\mu_n(\pi^{-1}_{\mathbf{R}^n \to \mathbf{R}^m}(E)) = \mu_m(E)$ for all Borel sets $E \subset \mathbf{R}^d$ by Exercise 11 of note 0. Since the spaces $(\mathbf{R}^n, \mathbf{B}[\mathbf{R}^n])$ are all standard Borel, by Theorem 10 there exists a unique probability measure $\mu_{\mathbf{N}}$ on $\mathbf{R}^{\mathbf{N}}$ satisfying the given property. $\square$

Exercise 14

Proof. Let G be the collection of Borel sets of $[0,1]^B$ on which μ_B is regular, then $G = \{E \in \mathcal{B}[[0,1]^B] : \forall \varepsilon > 0, \exists K \subset E \subset U \text{ such that } \mu_B(U \setminus K) < \varepsilon\}$, K compact and U open. First note that G contains all the open sets U. Clearly U can be approximated from above by itself. By Lemma 1.2.11 of "M", there is a sequence of closed boxes B_n such that $U = \bigcup_n B_n$, we can assume the sequence to be increasing(taking $B_n = \bigcup_n B_i$ if necessary), by continuity from below we get $\mu_B(U) = \lim_{n\to\infty} \mu_B(B_n)$, so U can be approximated from below by compact sets.

We claim that G is a Boolean algebra. Clearly $[0,1]^B \in G$ and $\emptyset \in G$. If $E_1, E_2 \in G$ and K_1, K_2, U_1, U_2 are such that K_1 and K_2 are compact sets approximating E_1 and E_2 respectively from below, and U_1 and U_2 are open sets approximating E_1 and E_2 respectively from above, then $K_1 \cup K_2$ is a compact set approximating $E_1 \cup E_2$ from below, and $U_1 \cup U_2$ is an open set approximating $E_1 \cup E_2$ from above, so $E_1 \cup E_2 \in G$. Similarly $E_1 \cap E_2 \in G$. By the monotone class lemma, $\langle G \rangle \subset \mathcal{B}[[0,1]^B]$. Conversely, as the Borel σ-algebra of $[0,1]^B$ is generated by the open sets, $\mathcal{B}[[0,1]^B] \subset \langle G \rangle$. Thus $\mathcal{B}[[0,1]^B] = \langle G \rangle$.

Let $P(E)$ be the property of a Borel set $E \subset [0,1]^B$ such that
$G = \{E \in \mathcal{B}[[0,1]^B] : P(E) \text{ is true}\}$. Assuming $P(E)$ is true, and $K \subset E \subset U$ are such that $\mu_B(U \setminus K) < \varepsilon$, K compact and U open, then $U^C \subset E^C \subset K^C$ are such that $\mu_B(K^C \setminus U^C) = \mu_B(U \setminus K) < \varepsilon$, U^C compact and K^C open. That is, $P(E^C)$ is true. Next, assuming $P(E_1), P(E_2), \ldots$ are all true, with $K_i \subset E_i \subset U_i$ be such that $\mu_B(U_i \setminus K_i) < \varepsilon/2^i$ for each i. Denote $\bigcup_{i=1}^{\infty} K_i$ by K and $\bigcup_{i=1}^{\infty} U_i$ by U, then $K \subset \bigcup_{i=1}^{\infty} E_i \subset U$ are such that $\mu_B(U \setminus K) < \varepsilon$. By continuity from below, $\mu_B(U \setminus K) = \lim_{n\to\infty} \mu_B(U \setminus \bigcup_{i=1}^{n} K_i) < \varepsilon$. Hence there

exists $n > 0$ such that $K' = \bigcup_{i=1}^n K_i$ and $\mu_B(U \setminus K') < \varepsilon$, K' compact and U open. i.e. $P(\bigcup_{i=1}^\infty E_i)$ is true. Clearly $P(E)$ is true for all $E \in G$, by Exercise 11 of Note 0 we get the desired result.

$\square$

Exercise 15

Proof. (1). This follows from the definition and the fact that $\mu_0(\emptyset) = 0$.

(2). This follows from the definition and the fact that $A \subset B$ implies $\inf A \leq \inf B$.

(3). Let $\epsilon > 0$. by the axiom of countable choice, we can choose an open elementary open covering $\bigcup_{i=1}^\infty E_{n,i}$ of E_n such that $\sum_{i=1}^\infty \mu_0(E_{n,i}) < \mu^*(E_n) + \epsilon/2^n$ for each n. From this, $\displaystyle\sum_{n=1}^\infty \sum_{i=1}^\infty \mu_0(E_{n,i}) < \sum_{n=1}^\infty \mu^*(E_n) + \epsilon$. Since $\epsilon > 0$ is arbitrary, LHS $\leq \sum_{n=1}^\infty \mu^*(E_n)$. As the double sequence $(E_{n,i})_{n,i}$ is an open elementary covering of $\bigcup_{n=1}^\infty E_n$, the Tonelli's theorem for series and the definition of the outer measure together give countable subadditivity.

(4). For any compact elementary set $E \subset [0,1]^A$, we claim that

$$\mu^*(E) = \inf_{V \supset E, \text{V open elementary}} \mu_0(V).$$

Let $E = \pi_B^{-1}(E_B)$ for some finite $B \subset A$ and some Borel $E_B \subset [0,1]^B$. Since π_B is continuous, E_B is compact. Note that by definition LHS $\leq$ RHS since every open elementary $V \supset E$ can be seen as a countable covering. Conversely, for every countable cover of E by open elementary sets E_i, $\pi_B(\bigcup_{i=1}^\infty E_i) = \bigcup_{i=1}^\infty \pi_B(E_i)$ is a countable cover of E_B by open sets(projection of open sets are open), and thus by Heine-Borel theorem there is a finite subcover $\bigcup_{i=1}^n \pi_B(E_i)$ of E_B. In particular, $V = \bigcup_{i=1}^n E_i$ is an open elementary cover of E such that

$$\sum_{i=1}^\infty \mu_0(E_i) \geq \sum_{i=1}^n \mu_0(E_i) \geq \mu_0(V) \geq \inf_{V \supset E, V \text{open elementary}} \mu_0(V).$$

This yields LHS $\geq$ RHS. Hence by outer regularity of μ_B we see that $\mu^*(E) = \mu_0(E)$ for compact elementary E.

To generalize to all elementary sets E, we show that the outer measure μ^* is inner regular on such E. That is, $\mu^*(E) = \sup_{K \subset E, K \text{compact}} \mu^*(K)$, from which follows that

$$\mu^*(E) = \sup_{K_B\subset E_B, K_B \text{compact}} \mu^*(\pi_B^{-1}(K_B)) = \sup_{K_B\subset E_B, K_B \text{compact}} \mu_0(\pi_B^{-1}(K_B)) =$$

$$\sup_{K_B\subset E_B, K_B \text{compact}} \mu_B(K_B) = \mu_B(E_B) = \mu_0(E) \text{ by inner regularity of } \mu_B.$$

By the regularity property of μ_B, we can find $K_B \subset E_B \subset U_B$, K_B compact and U_B open, with $K = \pi_B^{-1}(K_B)$ and $U = \pi_B^{-1}(U_B)$ such that $\mu^*(E) - \mu^*(K) \leq \mu_0(U) - \mu_0(K) = \mu_B(U_B) - \mu_B(K_B) \leq \varepsilon$ for any $\varepsilon > 0$, By monotonicity, this implies that the outer measure is inner regular on elementary sets, as desired.

(5). Let C be the collection of $E \in \mathcal{F}_A$ that can be approximated (in terms of outer measure) by elementary sets, clearly C is a Boolean algebra containing all elementary sets. By the monotone class lemma and the definition of $\mathcal{F}_A$, $\langle C\rangle = \mathcal{F}_A$. Now we use Exercise 11 of Note 0 to generalize the defining property of C to $\langle C\rangle$.

Let $E_1, E_2, \ldots \subset [0,1]^A$ be an increasing sequence in C (for any sequence E_n, $E = \bigcup_n E_n$ can be viewed as the countable union of the increasing sequence $A_n := \bigcup_{m=1}^n E_m$). It remains to show that $E \in C$. By construction, we can select an elementary sequence E'_n with $\mu^*(E_n \Delta E'_n) < \varepsilon/2^n$. Setting $E' = \bigcup_n E'_n$, by countable subadditivity we have: $\mu^*(E\Delta E') = \mu^*(\bigcup\limits_n E_n \Delta \bigcup\limits_n E'_n) \leq \mu^*(\bigcup\limits_n E_n \Delta E'_n) \leq \sum\limits_{n=1}^{\infty} \mu^*(E_n\Delta E'_n) < \sum\limits_{n=1}^{\infty} \varepsilon/2^n = \varepsilon$. On the other hand, let $F'_n = \bigcup_{m=1}^n E'_m$, because the bounded monotone sequence $\mu_0(F'_n)$ converges, $\mu^*(F'_n\backslash F'_{n-1}) = \mu_0(F'_n\backslash F'_{n-1})$ is summable ($F'_0 := \emptyset$). And their sum equals $\mu^*(E')$ by the definition of the outer measure. In particular, $\exists N > 0$ such that $\mu^*(E') - \sum_{n=1}^N \mu_0(F'_n\backslash F'_{n-1}) = \sum_{n=N+1}^{\infty} \mu_0(F'_n\backslash F'_{n-1}) < \varepsilon$. By countable subadditivity again, this means $\mu^*(E'\backslash F'_N) < \varepsilon$. By the triangle inequality, $\mu^*(E\Delta F'_N) \leq \mu^*(E\Delta E') + \mu^*(E'\backslash F'_N) < 2\varepsilon$, as desired. $\square$

Comment on Definition 16:

1. When talking about independence, all random variables under consideration are being modeled by a single probability space, such that the notion $\mathbf{P}(\bigwedge\limits_{i\in B} (X_i \in S_i)) = \mathbf{P}(\{\omega \in \Omega : X_i(\omega) \in S_i, i \in B\})$ makes sense.

2. To see why $\mathbf{P}(\bigwedge\limits_{j=1}^{J} (X_{A_j} \in S_j)) = \prod\limits_{j=1}^{J} \mathbf{P}(X_{A_j} \in S_j)$ follows from Exercise 3, let $B = \bigcup_{j=1}^J A_j$. By Exercise 3, $\prod\limits_{i\in B} \mu_{X_i} = \prod\limits_{j=1}^{J} (\prod\limits_{i\in A_j} \mu_{X_i})$. Therefore, writing $S_j = \prod_{i\in A_j} S_i$ for all j, we get: $\mathbf{P}(\bigwedge\limits_{j=1}^{J} (X_{A_j} \in S_j)) = \mathbf{P}(\bigwedge\limits_{i\in B} (X_i \in S_i)) =$

$\prod_{i\in B}\mathbf{P}(X_i\in S_i)=\prod_{i\in B}\mu_{X_i}(S_i)=\prod_{j=1}^{J}(\prod_{i\in A_j}\mu_{X_i}(S_i))=\prod_{j=1}^{J}\mathbf{P}(X_{A_j}\in S_j)$ by the fact that the X_i are jointly independent.

3. For the statement "This is the property of X and Y which is equivalent to independence (as can be seen by specialising to those F,G that take values in $\{0,1\}$)": If for instance $F=1_S$ and $G=1_T$ then $\mathbf{E}F(X)G(Y)=\mathbf{P}(X\in S\wedge Y\in T)$ and $\mathbf{E}F(X)\mathbf{E}G(Y)=\mathbf{P}(X\in S)\mathbf{P}(Y\in T)$.

4. For the reason why we fail to have independence in Example 17 if (X,Y) is drawn uniformly from say a parallelogram: If X,Y is drawn from a general shape (rather than a rectangle), then in general, knowing the value of X will give some information about Y and vice versa. This is for instance the observation that powers Berkson's paradox.

Exercise 18

Proof. Let X be a random variable taking values in a locally compact, σ-compact metric space R, and is independent of itself. Then for any measurable set $A\subset R$, $\mathbf{P}(X\in A)=\mathbf{P}(X\in A\wedge X\in A)=\mathbf{P}(X\in A)\mathbf{P}(X\in A)$, giving $\mathbf{P}(X\in A)\in\{0,1\}$. For any $x\in R$, the sequence of balls $B(x,1/n)$ is shrinking with n, hence either $\mathbf{P}(X\in B(x,1/n))=1$ for all n, or $\mathbf{P}(X\in B(x,1/n))=0$ for sufficiently large $n\geq N_x$. If the latter holds for all $x\in R$, then $B(x,1/N_x)$ forms an open cover of R, since every σ-compact space is Lindelöf, it induces a countable subcover $B(x_i,1/N_{x_i}),i=1,2,\ldots$ of R. By countable subadditivity, one would then get $1=\mathbf{P}(X\in R)\leq\sum_{i=1}^{\infty}\mathbf{P}(X\in B(x_i,1/N_{x_i}))=0$, a contradiction. Thus there exists at least one $x\in R$ such that $\mathbf{P}(X\in B(x,1/n))=1$ for all n, by continuity from above, $\mathbf{P}(X=x)=1$.

Conversely, if X is almost surely equal to a constant x, for all measurable sets $A,B\subset R$, $\mathbf{P}(X\in A\wedge X\in B)=\mathbf{P}(X\in A)\mathbf{P}(X\in B)$ holds trivially (depending on if $x\in A\cap B$ or not). That is, X is independent of itself. □

Exercise 19

Proof. Let X be a constant random variable, so $\mathbf{P}(X=c)=1$ for some constant c. For any measurable set A in the range of X, note that $\mathbf{P}(X\in A)=0$ (if $c\notin A$) or 1 (if $c\in A$). Hence for any other random variable Y and measurable set B in the range of Y, we have $\mathbf{P}(X\in A\wedge Y\in B)=\mathbf{P}(X\in A)\mathbf{P}(Y\in B)$ holds trivially. i.e. X is independent of any other random variable Y. □

Exercise 20

Proof. If $X_1, \dots, X_n$ are jointly independent, the conclusion follows from the definition of independence.

Conversely, suppose that $\mathbf{P}(\bigwedge_{i=1}^n (X_i = x_i)) = \prod_{i=1}^n \mathbf{P}(X_i = x_i)$ for all $x_1 \in R_1, \dots, x_n \in R_n$. Let $S_1 \subset R_1, \dots, S_n \subset R_n$, and note that $S = \prod_{i=1}^n S_i$ is at most countable since each S_i is. For clarification, write $X = (X_i)_{i \le 1 \le n}$, so $\mathbf{P}(\bigwedge\limits_{i=1}^n X_i \in S_i) = \mathbf{P}(X \in S) = \sum\limits_{x \in S} \mathbf{P}(X = x)$ by countable additivity.

On the other hand, we have $\prod_{i=1}^n \mathbf{P}(X_i \in S_i) = \prod_{i=1}^n (\sum_{x_i \in S_i} \mathbf{P}(X_i = x_i)) = \sum_{x \in S} \mathbf{P}(X = x)$ after expanding all the terms. Hence $\mathbf{P}(\bigwedge\limits_{i=1}^n X_i \in S_i) = \prod\limits_{i=1}^n \mathbf{P}(X_i \in S_i)$ and the X_i are independent of each other. □

Exercise 21

Proof. One direction follows trivially from the definition of independence.

Now suppose that $\mathbf{P}(\bigwedge_{i=1}^n (X_i \le t_i)) = \prod_{i=1}^n \mathbf{P}(X_i \le t_i)$ for all $t_1, \dots, t_n \in \mathbf{R}$, and let C be the collection of Borel subsets $S \subset \mathbf{R}^n$ such that $\mu_X(S) = (\prod_{i=1}^n \mu_{X_i})(S)$, where we denote the joint variable $(X_i)_{i=1}^n$ by X. By assumption C contains all boxes of the form $\prod_{i=1}^n (-\infty, t_i]$ with $t_i \in \mathbf{R}$. Now if $S_1 \subset S_2 \subset \dots$ is a countable increasing sequence in C, then:

$\mu_X(\bigcup\limits_{j=1}^\infty S_j) = \lim\limits_{j \to \infty} \mu_X(S_j) = \lim\limits_{j \to \infty} (\prod\limits_{i=1}^n \mu_{X_i})(S_j) = (\prod\limits_{i=1}^n \mu_{X_i})(\bigcup\limits_{j=1}^\infty S_j)$ by continuity from below. Thus $\bigcup_{j=1}^\infty S_j \in C$.

One can show the same for a decreasing sequence $S_1 \supset S_2 \supset \dots$ in C, we see that C is a monotone class. Now we show that C contains all half-open rectangular boxes of the form $\prod_{i=1}^n (a_i, b_i]$ with $a_i < b_i \in \mathbf{R}$, which follows from finite additivity:

$\mu_X(\prod\limits_{i=1}^n (a_i, b_i]) = \mathbf{P}(\bigwedge\limits_{i=1}^n (a_i < X_i \le b_i)) = \mathbf{P}(\bigwedge\limits_{i=1}^n (X_i \le b_i)) - \mathbf{P}(\bigwedge\limits_{i=1}^n (X_i \le a_i)) = \prod\limits_{i=1}^n \mathbf{P}(X_i \le b_i) - \prod\limits_{i=1}^n \mathbf{P}(X_i \le a_i) = \prod\limits_{i=1}^n \mathbf{P}(a_i < X_i \le b_i) = (\prod\limits_{i=1}^n \mu_{X_i})(\prod\limits_{i=1}^n (a_i, b_i])$.

Define a half-open elementary set to be a finite union of half-open rectangular boxes, and let A be the collection of all half-open elementary sets and their complements, we show that A is a Boolean algebra contained in C.

By construction, $\emptyset \in A$, and $S^C \in A$ whenever $S \in A$. If $S_1, S_2 \in A$, then $S_1 \cap S_2, S_1 \cup S_2 \in A$, since the union and intersection of two half open rectangular boxes B_1 and B_2 are both in A. For any $S \in A$, either $S = \bigcup_{j=1}^k B_j$, or $S = (\bigcup_{j=1}^k B_j)^C$, B_j half-open boxes, it suffice to show $S \in C$ in the former case. For convenience we assume the B_j's are disjoint, then again by finite additivity:

$$\mu_X(S) = \mu_X(\bigcup_{j=1}^{k} B_j) = \sum_{j=1}^{k} \mu_X(B_j) = \sum_{j=1}^{k}(\prod_{i=1}^{n} \mu_{X_i})(B_j) = (\prod_{i=1}^{n} \mu_{X_i})(\bigcup_{j=1}^{k} B_j) = (\prod_{i=1}^{n} \mu_{X_i})(S).$$

Finally, as boxes of the form $\prod_{i=1}^{n}(-\infty, t_i], t_i \in \mathbf{R}$ generate the Borel-σ algebra of $\mathbf{R}^n$, as can be seen by Exercise 1.4.14 of "M", A also generates the same algebra. By the monotone class theorem, we get the desired result. □

Exercise 22

Proof. Let $\beta = \{e_1, \ldots, e_r\}$ be an ordered basis of V, $|F| = q$. Since X is drawn uniformly at random from V, for any d-dimensional subspace S of V, $\mathbf{P}(X \in S) = q^d/q^r$.

WLOG, assume that $V = F^r$ (e.g. using the standard representation with respect to β). The conjunction of events $\bigwedge_{i=1}^{n}(\langle X, v_i\rangle = t_i)$ with $t_i \in F$ correspond to the system of linear equations $Bx = t$, where $B = \begin{pmatrix} \varphi(v_1) \\ \vdots \\ \varphi(v_n) \end{pmatrix} \in M_{n\times r}(F)$, $\varphi(v_i) = (\langle e_k, v_i\rangle)_{1\leq k\leq r}$ for all $1 \leq i \leq n$, $x \in V$ is viewed as a column vector, and $t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \in M_{n\times 1}(F)$. If s is the dimension of the span of $v_1, \ldots, v_n$, then $\operatorname{rank}(B) = s$ as B has exactly s independent rows $\varphi(v_i)$.(Note that non-degeneracy ensures we don't get 0 as rows as none of the $v_i = 0$). By Theorem 3.8 of "L", the dimension of the solution space equals $r - s$. It follows that $\mathbf{P}(\bigwedge_{i=1}^{n}(\langle X, v_i\rangle = t_i)) = q^{r-s}/q^r = 1/q^s$. Similarly, $\mathbf{P}(\langle X, v_i\rangle = t_i) = q^{r-1}/q^r = 1/q$. By Exercise 20, the random variables $\langle X, v_1\rangle, \ldots, \langle X, v_n\rangle$ are jointly independent if and only if $s = n$, i.e. if and only if the vectors $v_1, \ldots, v_n$ are linearly independent. □

Exercise 23

Consider the vector space $V = \mathbf{R}^2$ over a finite field F, and Y a random variable drawn uniformly at random from V. Let $\langle,\rangle : V \times V \to F$ be a non-degenerate bilinear form on V, and $u, v \in V$ not on the same line, $w := u + v$. Let $X, Y, Z := \langle Y, u\rangle, \langle Y, v\rangle, \langle Y, w\rangle$ respectively. Then we see by Exercise 22 that the random variables X, Y, Z are pair-wise independent but not jointly independent.

Exercise 24

Proof. First suppose the v_i are pairwise orthogonal, normalize the v_i such that $u_i = v_i/|v_i|$. By Theorem 6.7 of "L", we can extend the u_i into an orthonormal basis $\{u_i, 1 \leq i \leq n\}$ for $\mathbf{R}^n$.

Let U be the matrix with columns $u_1, \ldots, u_n$, then U is an orthogonal matrix by Exercise 23 in Section 6.1 of "L". Let $Y = U^T X$, and $g(x) = \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2}$ the pdf of the standard normal distribution on $\mathbf{R}^n$, by Exercise 1.3.16 of "M", the r.v's X and Y are equal in distribution:

$$\begin{aligned}\mathbf{P}(Y \in S) = \mathbf{P}(X \in US) &= \int_{\mathbf{R}^n} g 1_{US}(x)\ dx = \int_{\mathbf{R}^n} g 1_S(U^{-1}x)\ dx \\ &\overset{y=U^T x}{=} |\det(U)| \int_{\mathbf{R}^n} g(Uy) 1_S(y)\ dy = \int_S g(Uy)\ dy = \mathbf{P}(X \in S).\end{aligned}$$

That $|\det(U)| = 1$ follows from Exercise 12 in Section 4.3 of "L". For notational simplicity, write $X = (X_1, \ldots, X_n)$, and $g_i(x) = \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2}$ as the pdf of the standard normal distribution on $\mathbf{R}$. Since $g(x) = \prod_{i=1}^n g_i(x_i)$, we have by Fubini's theorem:

$$\begin{aligned}\mathbf{P}(\bigwedge_{i=1}^m (X \cdot v_i \in S_i)) = \mathbf{P}(\bigwedge_{i=1}^m (X \cdot u_i \in S_i/|v_i|)) &= \int_{\prod_{i=1}^n A_i} \prod_{i=1}^n g_i(x_i)\ d(x_1, \ldots, x_n) \\ &= \prod_{i=1}^n \int_{A_i} g_i(x_i)\ d(x_i) = (\prod_{i=1}^m \int_{S_i/|v_i|} g_i(x_i)\ d(x_i) \underbrace{(\prod_{i=m+1}^n \int_{\mathbf{R}} g_i(x_i)\ d(x_i))}_{=1} \\ &= \prod_{i=1}^m \int_{S_i/|v_i|} g_i(x_i)\ d(x_i) = \prod_{i=1}^m \mathbf{P}(X \cdot v_i \in S_i).\end{aligned}$$

Note the last equality follows again by rotational invariance of the standard Gaussian distribution, as $\mathbf{P}(X \cdot v_i \in S_i) = \mathbf{P}(UX \in \mathbf{R} \times \cdots \times S_i/|v_i| \times \cdots \times \mathbf{R}) = \int_{S_i/|v_i|} g_i(x_i)\ d(x_i)$. Hence the $X \cdot v_i$ are independent.

Conversely, assume $v_1, v_2 \in \mathbf{R}^n$ are not orthogonal. As $\mu_X = \prod_{i=1}^n \mu_{X_i}$ by condition, where μ_{X_i} is the standard normal distribution on $\mathbf{R}$, the components X_i are independent, and thus uncorrelated. Furthermore, either by direct computation or by Exercise 37 in note 1, $EX_i, E(X \cdot v_i) = 0$ for all $1 \leq i \leq n$, thus $EX_i X_j = 0$ for all $i \neq j$. By linearity:

$$\begin{aligned}\mathbf{E}(X \cdot v_1)(X \cdot v_2) &= \mathbf{E}(v_{1,1}X_1 + \ldots v_{1,n}X_n)(v_{2,1}X_1 + \ldots v_{2,n}X_n) = \sum_{i=1}^n v_{1,i} v_{2,i} \\ &\neq 0 = \mathbf{E}(X \cdot v_1)\mathbf{E}(X \cdot v_2).\end{aligned}$$

As $X \cdot v_1$ and $X \cdot v_2$ are not uncorrelated, they can't be independent. $\square$

Comment on equivalent formulation of a family of independent events: The cases where both 1 and 0 are in S_i and where none of 0 and 1 is in S_i are obvious. Hence we assume that for all finite $B \subset A$ and Borel subsets $S_i \subset \mathbf{R}, i \in B$, either $0 \in S_i$ or $1 \in S_i$, but not both. Undoing the definition, we have $\mathbf{P}(\bigwedge\limits_{i\in B}(1_{E_i} \in S_i)) = \mathbf{P}(\bigwedge\limits_{i\in A_1} E_i \wedge \bigwedge\limits_{j\in A_2} \overline{E_j}) = \prod\limits_{i\in A_1} \mathbf{P}(E_i) \prod\limits_{j\in A_2} \mathbf{P}(\overline{E_j})$, where $B = A_1 \uplus A_2$, $1 \in E_i, 0 \notin E_i$ for all $i \in A_1$ and $0 \in E_j, 1 \notin E_j$ for all $j \in A_2$.

Exercise 25

Proof. (1). By definition (and symmetry), E and F are jointly independent if and only if $\mathbf{P}(E\wedge\overline{F}) = \mathbf{P}(E)\mathbf{P}(\overline{F})$, which holds if and only if $\mathbf{P}(E)-\mathbf{P}(E\wedge F) = \mathbf{P}(E)(1-\mathbf{P}(F))$, if and only if $\mathbf{P}(E \wedge F) = \mathbf{P}(E)\mathbf{P}(F)$.

(2). The necessity of the condition follows from the definition of independent events. Now let $\Omega = \{H, T\}$ be the result of a fair coin toss, equipped with the discrete algebra. Let $E = \{H\}, F = \{T\}, G = \emptyset$, then $\mathbf{P}(E \wedge F \wedge G) = 0 = \mathbf{P}(E)\mathbf{P}(F)\mathbf{P}(G)$, but $\mathbf{P}((E \wedge F) \wedge \overline{G}) = 0 \neq 1/4 = \mathbf{P}(E)\mathbf{P}(F)\mathbf{P}(\overline{G})$. So the condition is not sufficient.

(3). Let $\Omega = \{1, 2, 3, 4\}$ with $\mathcal{F} = 2^\Omega$ and p the uniform distribution. If $E = \{1, 2\}, F = \{1, 3\}, G = \{2, 3\}$, then E, F, G are pairwise independent but not jointly independent. □

Exercise 27

Proof. (1). Let $X = \sum_{n=1}^\infty 2^{-n}\epsilon_n$, and $t \in [0, 1]$ with binary expansion $\sum\limits_{n=1}^{\infty} 2^{-n}t_n, t_n \in \{0, 1\}$. Note that $\mathbf{P}(\epsilon_n < t_n) = \frac{1}{2}t_n$ for all n, and $\mathbf{P}(X = t) = 0$. Hence:

$$\begin{aligned}
\mathbf{P}(X \le t) &= \mathbf{P}(X < t) + \mathbf{P}(X = t) \\
&= \mathbf{P}((\epsilon_1 < t_1) \vee (\epsilon_1 = t_1 \wedge \epsilon_2 < t_2) \vee (\epsilon_1 = t_1 \wedge \epsilon_2 = t_2 \wedge \epsilon_3 < t_3) \vee \dots) \\
&= \mathbf{P}(\epsilon_1 < t_1) + \mathbf{P}(\epsilon_1 = t_1 \wedge \epsilon_2 < t_2) + \mathbf{P}(\epsilon_1 = t_1 \wedge \epsilon_2 = t_2 \wedge \epsilon_3 < t_3) + \dots \\
&= \frac{1}{2}t_1 + (\frac{1}{2})^2 t_2 + (\frac{1}{2})^3 t_3 = t.
\end{aligned}$$

i.e. X is uniformly distributed on the unit interval $[0, 1]$.

(2). Let $X = \sum_{n=1}^{\infty} 3^{-n}(2\epsilon_n) = \sum_{n=1}^{\infty} 3^{-n} X_n$, $c : [0,1] \to [0,1]$ be the Cantor function, then X takes value in the middle thirds Cantor set C, and $\forall t \in C,\ t = \sum_{n=1}^{\infty} t_n 3^{-n}$, where $t_n \in \{0, 2\}$, by Exercise 1.6.47 of "M", we have

$$\begin{aligned}\mathbf{P}(X \le t) = \mathbf{P}(X < t) + \mathbf{P}(X = t) = \mathbf{P}(X < t) = \mathbf{P}((X_1 < t_1) \vee (X_1 = t_1 \wedge X_2 < t_2) \vee \dots) \\ = \mathbf{P}(X_1 < t_1) + \mathbf{P}(X_1 = t_1)\mathbf{P}(X_2 < t_2) + \dots = \frac{t_1}{2}2^{-1} + \frac{t_2}{2}2^{-2} + \dots \\ = c(t).\end{aligned}$$

Let $I = [a, b]$ be one of the intervals used in the n^{th} cover I_n of C, then $\mathbf{P}(X \in I) = \mathbf{P}(X \le b) - \mathbf{P}(X \le a) = |c(I)| = 2^{-n}$, again by Exercise 1.6.47 of "M". By the fact that the Cantor distribution is the unique probability distribution for which for any n ($n \in \{0, 1, 2, 3, \dots\}$), the probability of a particular interval in I_n containing the Cantor-distributed random variable is identically 2^{-n} on each one of the 2^n intervals, the claim holds. $\square$

Exercise 28

Example 1: Let $X \in \{0, 1\}$ be the Bernoulli r.v with expectation $1/2$, and $Y \in \{0, 1, -1\}$ be the real r.v such that $Y = 0$ when $X = 0$ and $Y = 1$ or -1 each with probability $1/2$ when $X = 1$. Then $E(XY) = E(X)E(Y) = 0$, i.e. the covariance $\mathrm{Cov}(X, Y) = 0$,but X and Y are dependent.

Example 2: Let X be a real r.v with the probability density function $x \mapsto \dfrac{1}{\sqrt{2\pi}}e^{-x^2/2}$ of the standard normal distribution, and $Y = X^2$. Then X and Y are dependent, but by Exercise 37 in note 0 the covariance $\mathrm{Cov}(X, Y) = 0$.

Exercise 29

Proof. For completeness we include the proof that the multivariate normal distribution indeed gives a pdf.

Denote $X := (X_1, \dots, X_n)$, By Theorem 6.20 of "L", Σ is orthogonally equivalent to a diagonal matrix D. Specifically, let $\beta = \{u_1, \dots, u_n\}$ be an orthonormal basis for $\mathbf{R}^n$ consisting of eigenvectors of Σ, let Q be the matrix whose columns are vectors in β, then Q is orthogonal and $D = Q^{-1}\Sigma Q$, or equivalently $\Sigma^{-1} = QD^{-1}Q^{-1}$, where D is the diagonal matrix with its j^{th} diagonal entry λ_j the eigenvalue of Σ corresponding to the j^{th} column of Q.

Fix a Borel subset $S \in \mathbf{R}^n$, and set $c = \dfrac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}}$. By linear change of variables, one obtains:

$$\begin{aligned}
\mathbf{P}(X \in \mathbf{R}^n) &\overset{y=x-\mu}{=} c\int_{\mathbf{R}^n} e^{-\frac{1}{2}y^T\Sigma^{-1}y}\ dy \\
&\overset{z=Q^Ty}{=} c|\det Q| \int_{\mathbf{R}^n} \exp(-\frac{1}{2}(Qz)^T\Sigma^{-1}(Qz))\ dz \\
&= c\int_{\mathbf{R}^n} e^{-\frac{1}{2}z^TD^{-1}z}\ dz \overset{w=z+\mu}{=} c\int_{\mathbf{R}^n} e^{-\frac{1}{2}(w-\mu)^TD^{-1}(w-\mu)}\ dw
\end{aligned}$$

The last integral splits by Fubini's theorem into:

$$c\int_{\mathbf{R}^n} \exp(-\frac{1}{2}\sum_{i=1}^{n}\frac{(w_i-\mu_i)^2}{\lambda_i})\ d(w_1,\dots,w_n) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\lambda_i}}\int_{\mathbf{R}}\exp(-\frac{1}{2}\frac{(w_i-\mu_i)^2}{\lambda_i})\ dw_i,$$

which equals 1. Note that the positive definite hypothesis ensures $\lambda_i > 0$ for all i (Exercise 17 in Section 6.4 of "L"), so the integrals on the RHS are all well defined.

(1). If $Y \in \mathbf{R}^n$ is the gaussian random vector with $Y \sim N(0_{\mathbf{R}^n}, D)$, then $Z = QY + \mu$ is a copy of X. Indeed, by condition,

$$\begin{aligned}
\mathbf{P}(X \in S) &= \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}}\int_S e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}\ dx \\
&\overset{y=Q^T(x-\mu)}{=} \frac{1}{(2\pi)^{n/2}(\det D)^{1/2}}\int_{Q^T(S-\mu)} e^{-\frac{1}{2}y^TD^{-1}y}\ dy \\
&= \mathbf{P}(Y \in Q^T(S-\mu)) = \mathbf{P}(QY+\mu \in S).
\end{aligned}$$

Taking the function f to be the ith coordinate projection π_i in Theorem 33 of note 1, we see that $\mathbf{E}X_i = \mathbf{E}Z_i$. Hence $\mathbf{E}X_i = \mathbf{E}(\sum_{j=1}^n Q_{ij}Y_j + \mu_i) = \sum_{j=1}^n Q_{ij}\mathbf{E}Y_j + \mu_i = \mu_i$ by linearity.

By the change of variables formula, for random vectors $U \overset{\mathrm{d}}{=} V$,

$\mathbf{E}(U_iU_j) = \int_{\mathbf{R}^n}\pi_1(u)\pi_2(u)\ d\mu_U(u) = \int_{\mathbf{R}^n}\pi_1(v)\pi_2(v)\ d\mu_V(v) = \mathbf{E}(V_iV_j)$, so in particular $\mathrm{Cov}(U_i,U_j) = \mathrm{Cov}(V_i,V_j)$. Since the components of Y are independent, we get:

$$\mathrm{Cov}(X_i,X_j) = \mathrm{Cov}(Z_i,Z_j) = \sum_{r=1}^{n} Q_{ir}Q_{jr}\mathbf{E}(Y_r^2) = \sum_{r=1}^{n}\lambda_rQ_{ir}Q_{jr} = \sigma_{ij}.$$

We haven shown that every positive definite real symmetric matrix Σ is the covariance matrix of some gaussian random vector $Z = QY + \mu$.

(2). By definition, independence implies vanishing covariance. Conversely, if all the covariances $\mathrm{Cov}(X_i,X_j)$ for $1 \le i \ne j \le n$ vanish, Σ becomes a diagonal matrix $diag(\sigma_{11},\dots,\sigma_{nn})$. Hence the pdf of X splits and we have:

$\mathbf{P}(X \in \prod_{i=1}^{n} S_i) = \prod_{i=1}^{n}\int_{S_i}\exp(-\frac{1}{2}\frac{(x_i-\mu_i)^2}{\sigma_{ii}})\ dx_i = \prod_{i=1}^{n}\mathbf{P}(X_i \in S_i)$. That is, $(X_1,\dots,X_n)$ are jointly independent.

(3). Let X be a real standard gaussian r.v, and R a random variable taking 1 and -1 each with probability 1/2 (a random sign). Then $Y = RX$ is such that: $\mathbf{P}(Y \in S) = \mathbf{P}(RX \in S) = \mathbf{P}(R = 1)\mathbf{P}(X \in S) + \mathbf{P}(R = -1)\mathbf{P}(X \in -S) = \dfrac{1}{2\sqrt{2\pi}} \displaystyle\int_S e^{-\frac{1}{2}x^2}\, dx + \dfrac{1}{2\sqrt{2\pi}} \displaystyle\int_{-S} e^{-\frac{1}{2}x^2}\, dx = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_S e^{-\frac{1}{2}x^2}\, dx.$

i.e. Y is also standard gaussian. Clearly X and Y are not independent, but $\mathrm{Cov}(X, Y) = \mathbf{E}(RX^2) - \mathbf{E}(X)\mathbf{E}(RX) = \mathbf{E}(R)[\mathbf{E}(X^2) - \mathbf{E}(X^2)] = 0.$ □

Exercise 31

Proof. Let $B \subset A$ be finite. The event $E_i \in \sigma(X_i)$ for any $i \in B$ corresponds with the event $X_i \in S_i$ for some measurable set S_i in the range of X_i, hence by definition $(X_i)_{i \in A}$ are jointly independent random variables if and only if $(\sigma(X_i))_{i \in A}$ are jointly independent σ-algebras. □

Exercise 32

Proof. Let $(R_n, \mathcal{B}_n)$ be the range of X_n for $n = 1, 2, \ldots$ For sufficiency, assume that $\sigma(X_{n+1})$ is independent of $\sigma(X_1, \ldots, X_n)$ for all natural numbers n. Let $B \subset \mathbf{N}$ be finite with $|B| > 1$ (the case for $|B| = 1$ being trivial). Let $N := \max(B)$, if we set $S'_n := S_n$ for $n \in B$ and $S'_n := R_n$ otherwise, then:

$$\begin{aligned}\mathbf{P}(\bigwedge_{n \in B} (X_n \in S_n)) &= \mathbf{P}(\bigwedge_{n=1}^{N} (X_n \in S'_n)) = \mathbf{P}(\bigwedge_{n=1}^{N-1} (X_n \in S'_n) \wedge (X_N \in S'_N)) \\ &= \mathbf{P}(E_1 \wedge E_2) = \mathbf{P}(E_1)\mathbf{P}(E_2) \\ &= \mathbf{P}(\bigwedge_{n=1}^{N-1} (X_n \in S'_n))\mathbf{P}(X_N \in S'_N).\end{aligned}$$

Where $E_1 \in \sigma(X_1, \ldots, X_{N-1}),\ E_2 \in \sigma(N)$. By iteration, one gets $\mathbf{P}(\bigwedge_{n \in B} X_n \in S_n) = \prod_{n=1}^{N} \mathbf{P}(X_n \in S'_n) = \prod_{n \in B} \mathbf{P}(X_n \in S_n)$. That is, $(X_n)_{n=1}^{\infty}$ are jointly independent.

Conversely, suppose that $(X_n)_{n=1}^{\infty}$ are jointly independent. $\forall n \in \mathbf{N}$, by definition, $\sigma(X_1, \ldots, X_n) = \langle \bigcup_{i=1}^{n} \sigma(X_i) \rangle$, and by the previous Exercise $\sigma(X_{n+1})$ is independent of $\sigma(X_i)$ for all $1 \leq i \leq n$. We use induction to extend from the the set of generators to the entire σ-algebra.

We let $P(E)$ be the property of any event $E \subset \Omega$ that $\mathbf{P}(E \cap F) = \mathbf{P}(E)\mathbf{P}(F)$ for any event $F \in \sigma(X_{n+1})$. $P(\emptyset)$ holds trivially since any event is independent of the empty event. If $\mathbf{P}(E \cap F) = \mathbf{P}(E)\mathbf{P}(F)$, then $\mathbf{P}(E^C \cap F) = \mathbf{P}(F \setminus E) = \mathbf{P}(F) - \mathbf{P}(E \cap F) = \mathbf{P}(F) - \mathbf{P}(E)\mathbf{P}(F) = \mathbf{P}(E^C)\mathbf{P}(F)$ by finite additivity.

Lastly, If $E_1, E_2, \dots \subset \Omega$ are such that $P(E_n)$ is true for all n, without loss of generality we assume that the E_n are pairwise disjoint, then $\mathbf{P}((\bigcup_{n=1}^{\infty} E_n) \cap F) = \mathbf{P}(\bigcup_{n=1}^{\infty}(E_n \cap F)) = \sum_{n=1}^{\infty} \mathbf{P}(E_n \cap F) = \sum_{n=1}^{\infty} \mathbf{P}(E_n)\mathbf{P}(F) = \mathbf{P}(\bigcup_n E_n)\mathbf{P}(F)$ by countable additivity. The claim then follows by induction. $\square$

Comment on the proof of Theorem 33 (Kolmogorov zero-one law):

1. The algebras $\sigma(X_1, \dots, X_n)$ are increasing in n, by Exercise 32 and the fact that independent to the larger algebra implies independent to the smaller algebra, the generators $\sigma(X_{n+1}), \sigma(X_{n+2}), \dots$ of $\sigma(X_i : i > n)$ are all independent to $\sigma(X_1, \dots, X_n)$, by induction the σ-algebra $\sigma(X_i : i > n)$ is thus independent of $\sigma(X_1, \dots, X_n)$.

2. Let $C = \{E \in \sigma(X_i : i \geq 1) : \text{E is independent to all tail events } F \in \mathcal{T}\}$, then C is a monotone class containing $\sigma(X_1, \dots, X_n)$ for all $n = 1, 2, \dots$, by monotone class lemma $C \supset \sigma(X_i : i \geq 1)$. i.e. $\mathcal{T}$ is independent of $\sigma(X_i : i \geq 1)$.