Solutions

Chapter 2 Foundations of Probability

2.1 (Composing random elements) Show that if f is $\mathcal{F}/\mathcal{G}$-measurable and g is $\mathcal{G}/\mathcal{H}$-measurable for sigma algebras $\mathcal{F}$,$\mathcal{G}$ and $\mathcal{H}$ over appropriate spaces, then their composition, $g \circ f$ (defined the usual way: $(g \circ f)(\omega) = g(f(\omega)), \omega \in \Omega$), is $\mathcal{F}/\mathcal{H}$-measurable.

Proof. Since g is $\mathcal{G}/\mathcal{H}$-measurable, therefore $\forall C \in \mathcal{H}$, $\exists B = g^{-1}(C) \in \mathcal{G}$. Similarly, since f is $\mathcal{F}/\mathcal{G}$-measurable, $\forall B \in \mathcal{G}$, $\exists A = f^{-1}(B) \in \mathcal{F}$. Thus $\forall C \in \mathcal{H}$, $\exists A = f^{-1}(g^{-1}(C)) = (g \circ f)^{-1}(C) \in \mathcal{F}$ and the proof is complete.

□

2.2 Let $X_1, \ldots, X_n$ be random variables on $(\Omega, \mathcal{F})$. Prove that $X = (X_1, \ldots, X_n)$ is a random vector.

Proof. Since X_i is a random variable ($\forall i = 1, 2, ..., n$), it holds that X_i is $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable, which means that $\forall B \in \mathcal{B}(\mathbb{R})$, $X_i^{-1}(B) \in \mathcal{F}$. We first prove that X is $\mathcal{F}/(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \times \cdots \mathcal{B}(\mathbb{R}))$-measurable (totally n $\mathcal{B}(\mathbb{R})$s). $\forall A = A_1 \times A_2 \times \cdots \times A_n \in \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \times \cdots \mathcal{B}(\mathbb{R})$, $X^{-1}(A) = X_1^{-1}(A_1) \cap X_2^{-1}(A_2) \cap \cdots \cap X_n^{-1}(A_n) \in \mathcal{F}$, which holds since $X_i^{-1}(A_i) \in \mathcal{F}, \forall i = 1, 2, ..., n$ and $\mathcal{F}$ is a σ-algebra. Thus we conclude that X is $\mathcal{F}/(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \times \cdots \mathcal{B}(\mathbb{R}))$-measurable.

By definition $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \times \cdots \mathcal{B}(\mathbb{R}))$ (totally n $\mathcal{B}(\mathbb{R})$s). And according to the property in 2.5(b), we can get that X is $\mathcal{F}/\mathcal{B}(\mathbb{R}^n)$-measurable, thus it is a random vector.

□

2.3 (Random variable induced σ-algebra) Let $\mathcal{U}$ be an arbitrary set and $(\mathcal{V}, \Sigma)$ a measurable space and $X : \mathcal{U} \to \mathcal{V}$ an arbitrary function. Show that $\Sigma_X = \{X^{-1}(A) : A \in \Sigma\}$ is a σ-algebra over $\mathcal{U}$.

Proof.

(i) We need to show that Σ_X is closed under countable union. Let $U_i = X^{-1}(A_i), A_i \in \Sigma, i \in \mathbb{N}$. It follows that $\bigcup\limits_{i=1}^{\infty} U_i = \bigcup\limits_{i=1}^{\infty} X^{-1}(A_i) = X^{-1}(\bigcup\limits_{i=1}^{\infty} A_i)$. Since $\bigcup\limits_{i=1}^{\infty} A_i \in \Sigma$, $\bigcup\limits_{i=1}^{\infty} U_i \in \Sigma_X$.

(ii) We need to show that Σ_X is closed under set subtraction $-$. $\forall U_1, U_2 \in \Sigma_X$,$U_1 - U_2 = X^{-1}(A_1) - X^{-1}(A_2) = X^{-1}(A_1 - A_2)$. Since $A_1 - A_2 \in \Sigma$, $U_1 - U_2 \in \Sigma_X$.

(iii) We need to show that Σ_X is closed to $\mathcal{U}$ itself. Since $\mathcal{U} = X^{-1}(\mathcal{V})$ and $\mathcal{V} \in \Sigma$, it follows that $\mathcal{U} \in \Sigma_X$.

□

2.4 Let $(\Omega, \mathcal{F})$ be a measurable space and $A \subseteq \Omega$ and $\mathcal{F}_{|A} = \{A \cap B : B \in \mathcal{F}\}$.

Proof. (a)

(i) We need to show that $\mathcal{F}|_A$ is closed under countable union. Let $X_1 = A \cap B_1, X_2 = A \cap B_2, ...$ and $X' = \bigcup\limits_{i=1}^{\infty} X_i$ and $B' = \bigcup\limits_{i=1}^{\infty} B_i$ where $B_1, B_2, ... \in \mathcal{F}$. Since $\mathcal{F}$ is sigma algebra, $B' \in \mathcal{F}$. Furthermore, since $X' = \bigcup_{i=1}^{\infty} X_i = \bigcup_{i=1}^{\infty} A \bigcap B_i = A \bigcap \left(\bigcup\limits_{i=1}^{\infty} B_i \right) = A \bigcap B'$, we can see that $X' \in \mathcal{F}|_A$.

(ii) We need to show that $\mathcal{F}|_A$ is closed under set subtraction $-$. $\forall X_1, X_2 \in \mathcal{F}|_A$, $X_1 - X_2 = (A \bigcap B_1) - (A \bigcap B_2) = A \bigcap (B_1 - B_2)$. Since $B_1 - B_2 \in \mathcal{F}$, it follows that $X_1 - X_2 \in \mathcal{F}|_A$.

(iii) We need to show that $\mathcal{F}|_A$ is closed to A itself. Since $\varnothing \in \mathcal{F}$, we have $\varnothing = A \bigcap \varnothing \in \mathcal{F}|_A$ and $A = \varnothing^C \in \mathcal{F}|_A$.

(b) Let $P = \{A \bigcap B : B \in \mathcal{F}\}, Q = \{B : B \subset A, B \in \mathcal{F}\}$.

(i) We claim that $P \subset Q$. Let $X = A \bigcap B, B \in \mathcal{F}$. Since $A \in \mathcal{F}$, $X = A \bigcap B \in \mathcal{F}$. Furthermore, $X \in Q = \{B : B \subset A, B \in \mathcal{F}\}$.

(ii) We claim that $Q \subset P$. $\forall X \in Q$, we have $X \subset A$ and $X \in \mathcal{F}$, which means that $X = X \bigcap A$ and $X \in \mathcal{F}$. It follows that $X \in P$.

(iii) Take both (i)(ii) into consideration, we can see that $P = Q$.

□

2.5 Let $\mathcal{G} \subseteq 2^\Omega$ be a non-empty collection of sets and define $\sigma(\mathcal{G})$ as the smallest σ-algebra that contains $\mathcal{G}$. By 'smallest' we mean that $\mathcal{F} \in 2^\Omega$ is smaller than $\mathcal{F}' \in 2^\Omega$ if $\mathcal{F} \subset \mathcal{F}'$.

(a) Show that $\sigma(\mathcal{G})$ exists and contains exactly those sets A that are in every σ-algebra that contains $\mathcal{G}$.

(b) Suppose $(\Omega', \mathcal{F})$ is a measurable space and $X : \Omega' \to \Omega$ be $\mathcal{F}/\mathcal{G}$-measurable. Show that X is also $\mathcal{F}/\sigma(\mathcal{G})$-measurable. (We often use this result to simplify the job of checking whether a random variable satisfies some measurability property).

(c) Prove that if $A \in \mathcal{F}$ where $\mathcal{F}$ is a σ-algebra, then $\mathbb{I}\{A\}$ is $\mathcal{F}$-measurable.

Proof. (a) Let $\mathcal{K} = \{\mathcal{F} | \mathcal{F}$ is a σ-algebra and contains $\mathcal{G}\}$, It holds obviously that $\mathcal{K}$ is not an empty set since it contains $2^\mathcal{G}$.

Then $\bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$ contains exactly those sets that are in every σ-algebra that contains $\mathcal{G}$. Given its existence, we only need to prove that $\bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$ is the smallest σ-algebra that contains $\mathcal{G}$.

First we show $\bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$ is a σ-algebra. Since $\mathcal{F}$ is a σ-algebra and therefore $\Omega \in \mathcal{F}$ for all $\mathcal{F} \in \mathcal{K}$, it follows that $\Omega \in \bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$. Next, for any $A \in \bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$, $A^c \in \mathcal{F}$ for all $\mathcal{F} \in \mathcal{K}$. Since they are all σ-algebras, $A^c \in \mathcal{F}$ for all $\mathcal{F} \in \mathcal{K}$. Hence $A^c \in \bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$. Finally, for any $\{A_i\}_i \subset \bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$, $\{A_i\}_i \subset \mathcal{F}$ for all $\mathcal{F} \in \mathcal{K}$. Since they are all σ-algebras, $\bigcup_i A_i \in \mathcal{F}$ for all $\mathcal{F} \in \mathcal{K}$. Hence $\bigcup_i A_i \in \bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$.

Next we want to prove $\bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$ is the smallest σ-algebra that contains $\mathcal{G}$. It is quite obvious that $\bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F} \subseteq \mathcal{F}'$ for all $\mathcal{F}' \in \mathcal{K}$.

Above all, we have $\sigma(\mathcal{G}) = \bigcap_{\mathcal{F} \in \mathcal{K}} \mathcal{F}$.

(b) Define $\mathcal{H} = \{A : X^{-1}(A) \in \mathcal{F}\}$. To show X is $\mathcal{F}/\sigma(\mathcal{G})$-measurable, it is sufficient to prove $\sigma(\mathcal{G}) \subseteq \mathcal{H}$.

First we prove that $\mathcal{H}$ is a σ-algebra. It holds that $\Omega \in \mathcal{H}$ since $X^{-1}(\Omega) = \Omega' \in \mathcal{F}$. For any $A \in \mathcal{H}$, we have $X^{-1}(A) \in \mathcal{F}$, thus $X^{-1}(A^c) = X^{-1}(A)^c \in \mathcal{F}$, which holds since $\mathcal{F}$ is a σ-algebra. Thus $A^c \in \mathcal{H}$. For any $A_i \in \mathcal{F}$, $i = 1, 2, ...$, $X^{-1}(A_i) \in \mathcal{F}$, $X^{-1}(\cup_i A_i) = \cup_i X^{-1}(A_i) \in \mathcal{F}$. We can then conclude $\cup_i A_i \in \mathcal{H}$ and $\mathcal{H}$ is a σ-algebra.

Also, since X is $\mathcal{F}/\mathcal{G}$-measurable, we have $\mathcal{G} \subseteq \mathcal{H}$. Thus $\mathcal{H}$ is σ-algebra that contains $\mathcal{G}$. By applying the result of (a), we have $\sigma(\mathcal{G}) \subseteq \mathcal{H}$, which completes the proof.

(c) The idea is to show $\forall B \in \mathfrak{B}(\mathbb{R})$, $\mathbb{I}\{A\}^{-1}(B) \in \mathcal{F}$.

If $\{0, 1\} \in B$, $\mathbb{I}\{A\}^{-1}(B) = \Omega \in \mathcal{F}$. If $\{0\} \in B$, $\mathbb{I}\{A\}^{-1}(B) = A^c \in \mathcal{F}$. If $\{1\} \in B$, $\mathbb{I}\{A\}^{-1}(B) = A \in \mathcal{F}$. If $\{0, 1\} \cap B = \emptyset$, $\mathbb{I}\{A\}^{-1}(B) = \emptyset \in \mathcal{F}$.

□

2.6 (Knowledge and σ-algebras: a pathological example) In the context of Lemma 2.5, show an example where $Y = X$ and yet Y is not $\sigma(X)$ measurable.

Hint As suggested after the lemma, this can be arranged by choosing $\Omega = \mathcal{Y} = \mathcal{X} = \mathbb{R}, X(\omega) = Y(\omega) = \omega, \mathcal{F} = \mathcal{H} = \mathfrak{B}(\mathbb{R})$ and $\mathcal{G} = \{\emptyset, \mathbb{R}\}$ to be the trivial σ-algebra.

Proof. As the hint suggests, Let $\Omega = \mathcal{Y} = \mathcal{X} = \mathbb{R}, X(\omega) = Y(\omega) = \omega, \mathcal{F} = \mathcal{H} = \mathfrak{B}(\mathbb{R})$. In this case, $\sigma(X) = \{X^{-1}(A) : A \in \mathcal{G}\} = \{\emptyset, \mathbb{R}\}$, we can find that $Y^{-1}((0,1)) = (0,1) \notin \sigma(X)$, thus Y is not $\sigma(X)$-measurable. □

2.7 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $B \in \mathcal{F}$ be such that $\mathbb{P}(B) > 0$. Prove that $A \mapsto \mathbb{P}(A|B)$ is a probability measure over $(\Omega, \mathcal{F})$.

Proof. First we have $\mathbb{P}(\Omega \mid B) = \frac{\mathbb{P}(\Omega\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$. Then, for any $A \in \mathcal{F}$, $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \geq 0$. Next, for any $A \in \mathcal{F}$, $\mathbb{P}(A^c \mid B) = \frac{\mathbb{P}(A^c\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}((\Omega - A)\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B) - \mathbb{P}(A\cap B)}{\mathbb{P}(B)} = 1 - \mathbb{P}(A \mid B)$. Finally, for all countable collections of disjoint sets $\{A_i\}_i$ with $A_i \in \mathcal{F}$ for all i, we have $\mathbb{P}\left(\bigcup_i A_i \mid B\right) = \frac{\mathbb{P}((\bigcup_i A_i)\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\bigcup_i (A_i \cap B))}{\mathbb{P}(B)} = \sum_i \frac{\mathbb{P}(A_i\cap B)}{\mathbb{P}(B)} = \sum_i \mathbb{P}(A_i \mid B)$. □

2.8 (Bayes law) Verify (2.2).

Proof. With the definition of conditional probability, we have $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$. □

2.9 Consider the standard probability space $(\Omega, \mathcal{F}, \mathbb{P})$ generated by two standard, unbiased, six-sided dice that are thrown independently of each other. Thus, $\Omega = \{1, ..., 6\}^2$, $\mathcal{F} = 2^\Omega$ and $\mathbb{P}(A) = |A|/6^2$ for any $A \in \mathcal{F}$ so that $X_i(\omega) = \omega_i$ represents the outcome of throwing dice $i \in \{1, 2\}$.

(a) Show that the events '$X_1 < 2$' and 'X_2 is even' are independent of each other.

(b) More generally, show that for any two events, $A \in \sigma(X_1)$ and $B \in \sigma(X_2)$, are independent of each other.

Proof. (a) The event $\{X_1 < 2\} = \{1\} \times \{1,2,3,4,5,6\}$, $\{X_2 \text{ is even }\} = \{1,2,3,4,5,6\} \times \{2,4,6\}$, $\{X_1 < 2, X_2 \text{ is even }\} = \{(1,2),(1,4),(1,6)\}$.

Thus $\mathbb{P}(X_1 < 2) = \frac{6}{36} = \frac{1}{6}$, $\mathbb{P}(X_2 \text{ is even }) = \frac{18}{36} = \frac{1}{2}$, $\mathbb{P}(X_1 < 2, X_2 \text{ is even }) = \frac{3}{36} = \frac{1}{12}$, which satisfies $\mathbb{P}(X_1 < 2, X_2 \text{ is even }) = \mathbb{P}(X_1 < 2) \times \mathbb{P}(X_2 \text{ is even })$. These two events are independent of each other.

(b) $\sigma(X_1) = \left\{X_1^{-1}(A'), A' \subseteq [6]\right\} = \{A' \times [6] : A' \subseteq [6]\}$, $\sigma(X_2) = \left\{X_2^{-1}(B'), B' \subseteq [6]\right\} = \{[6] \times B' : B' \subseteq [6]\}$.
Thus $\forall A \in \sigma(X_1), B \in \sigma(X_2)$, $\mathbb{P}(A) = \frac{|A'|\times 6}{36} = \frac{|A'|}{6}$, $\mathbb{P}(B) = \frac{6\times|B'|}{36} = \frac{|B'|}{6}$ and $\mathbb{P}(A \cap B) = \frac{|A'|\times|B'|}{36} = \mathbb{P}(A) \times \mathbb{P}(B)$. So A and B are independent of each other. □

2.10 (Serendipitous independence) The point of this exercise is to understand independence more deeply. Solve the following problems:

(a) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Show that $\emptyset$ and Ω (which are events) are independent of any other event. What is the intuitive meaning of this?

(b) Continuing the previous part, show that any event $A \in \mathcal{F}$ with $\mathbb{P}(A) \in \{0, 1\}$ is independent of any other event.

(c) What can we conclude about an event $A \in \mathcal{F}$ that is independent of its complement, $A^c = \Omega \setminus A$? Does your conclusion make intuitive sense?

(d) What can we conclude about an event $A \in \mathcal{F}$ that is independent of itself? Does your conclusion make intuitive sense?

(e) Consider the probability space generated by two independent flips of unbiased coins with the smallest possible σ-algebra. Enumerate all pairs of events A, B such that A and B are independent of each other.

(f) Consider the probability space generated by the independent rolls of two unbiased three-sided dice. Call the possible outcomes of the individual dice rolls 1, 2 and 3. Let X_i be the random variable that corresponds to the outcome of the ith dice roll ($i \in \{1, 2\}$). Show that the events $\{X_1 \le 2\}$ and $\{X_1 = X_2\}$ are independent of each other.

(g) The probability space of the previous example is an example when the probability measure is uniform on a finite outcome space (which happens to have a product structure). Now consider any n-element, finite outcome space with the uniform measure. Show that A and B are independent of each other if and only if the cardinalities $|A|$, $|B|$, $|A \cap B|$ satisfy $n|A \cap B| = |A| \cdot |B|$.

(h) Continuing with the previous problem, show that if n is prime, then no non-trivial events are independent (an event A is **trivial** if $\mathbb{P}(A) \in \{0, 1\}$).

(i) Construct an example showing that pairwise independence does not imply mutual independence.

(j) Is it true or not that A, B, C are mutually independent if and only if $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$? Prove your claim.

Proof. (a) Empty sets and complete sets are independent of any event:

$$P(A \cap \Omega) = P(A) = 1 \times P(A) = P(\Omega) \times P(A)$$
$$P(A \cap \emptyset) = P(\emptyset) = 0 = P(\emptyset) \times P(A)$$

(b) For any $B \in \Omega$ and $P(A) \in \{0, 1\}$:
when $P(A) = 1$,$P(A^c \cap B) \le P(A^c) = 1 - P(A) = 0$, we have $P(A \cap B) = P(A \cap B) + P(A^c \cap B) = P(B) = P(A)P(B)$; when $P(A) = 0$,we have $P(A \cap B) \le P(A) = 0 = P(A)P(B)$

(c) $P(A^c \cap A) = P(A)P(A^c)$, we have $0 = P(A)(1 - P(A)) \Rightarrow P(A) \in \{0, 1\}$

(d) $P(A \cap A) = P(A)P(A)$, we have $P(A) = \{0, 1\}$

(e) $\Omega = \{(1, 1), (1, 0), (0, 1), (0, 0)\}.A, B \subseteq \Omega$ denote the events.

First of all, if either A or B is trival, then A and B are independent of each other.

Then, we only need to enumerate $A, B \notin \Omega, \emptyset$ satisfied that $P(A \bigcap B) = P(A)P(B)$. Since $P(A \bigcap B) = \frac{|A \bigcap B|}{|\Omega|} = \frac{|A \bigcap B|}{4}$ and $P(A)P(B) = \frac{|A||B|}{16}$, we can conclude that $|A| = 2$, $|B| = 2$ and $|A \bigcap B| = 1$ is the only situation satisfying the condition.

Thus, besides trival A or B, all A, B satisfying $|A| = 2$, $|B| = 2$ and $|A \bigcap B| = 1$ are the solution.

(f) $P(X_1 \le 2) = 2/3$

$P(X_1 = X_2) = 3/9 = 1/3$

$P(X_1 \le 2, X_1 = X_2) = P(X_1 = X_2 = 1) + P(X_1 = X_2 = 2) = 1/9 + 1/9 = 2/9$

So, $P(X_1 \le 2, X_1 = X_2) = P(X_1 = X_2)P(X_1 \le 2)$

(g) Necessity :$\frac{|A \bigcap B|}{n} = P(A \bigcap B) = P(A)P(B) = \frac{|A|}{n}\frac{|B|}{n}$

$\Rightarrow |A \bigcap B| \times n = |A||B|$

Sufficiency :$|A \bigcap B| \times n = |A||B| \Rightarrow \frac{|A|}{n}\frac{|B|}{n} = \frac{|A \bigcap B|}{n}$

$\Rightarrow P(A \bigcap B) = P(A)P(B)$

(h) If A, B are two non-trival events independent to each other, $|A \bigcap B| \times n = |A||B| \Rightarrow n|\ (|A||B|) \Rightarrow n|\ (|A|)$ or $n|\ (|B|) \Rightarrow |A| = n$ or $|B| = n$, contradictory to non-trival assumption.

(i) Let $\Omega = \{1, 2, 3, 4\}$, $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$. A, B, C are pairwise independent but $P(A \bigcap B \bigcap C) = \frac{1}{4} \neq P(A)P(B)P(C) = \frac{1}{8}$.

(j) Consider rolling a dice and set $A = \{1,2,3\}$, $B = \{1,2,4\}$,$C = \{1,4,5,6\}$. Then $P(A \cap B \cap C) = \frac{1}{6} = (1/2) * (1/2) * (2/3) = P(A)P(B)P(C)$, however $P(A \cap B) = 1/3 \neq \frac{1}{2} * \frac{1}{2} = P(A)P(B)$. Thus $P(A \cap B \cap C) = P(A)P(B)P(C)$ does not mean mutuall independence.

$\square$

2.11

(a) X:$\Omega \to x$

Because X, Y are independent equivalent to$\sigma(X)$,$\sigma(Y)$ are independent;

For any A$\in \sigma(Y)$,

$$P(\phi \bigcap A) = P(\phi) = 0 = P(\phi)P(A)$$

$$P(\Omega \bigcap A) = P(A) = P(\Omega)P(A)$$

(b) We know that $P(X = x) = 1$

$$P(X = x|Y) = \frac{P((X = x)\bigcap Y)}{P(Y)} = 1 = P(X = x)$$

$$P(X \neq x|Y) = 1 - P(X = x|Y) = 0 = P(X \neq x)$$

(c) Notice the relation:$P(A) = P(X(A) = 1)$

$P(B) = P(X(B) = 1)$

$P(A \bigcap B) = P(X(A \bigcap B) = 1)$

The first two formulas follow the definition. Let's prove the third equation:

$P(X(A \bigcap B) = P(X(A) + X(B) - X(A \bigcup B) = 1)$

Let's discuss$X(A), X(B), X(A \bigcup B)$:

X(A)	X(B)	X(A$\bigcup$ B)	$X(A) + X(B) - X(A \bigcup B)$
1	1	1	1
1	0	1	0
0	1	1	0
0	0	0	0

We can see that,$P(X(A \bigcap B) = P(X(A) + X(B) - X(A \bigcup B) = 1)$, this is only one case of the first row of the table.

that is$P(X(A \bigcap B) = 1) = P(X(A) = 1, X(B) = 1) = P(A \bigcap B)$

that is$P(X(A \bigcap B) = 1) = P(A \bigcap B)$

So,$P(A \bigcap B) = P(A)P(B)$ is equivalent to $P(X(A \bigcap B) = 1) = P(X(A) = 1)P(X(B) = 1)$

(d) A_i pairwise i $\Leftrightarrow I\{A_i\}$ pairwise i

mutual i $\Leftrightarrow P(\bigcap_i A_i) = \prod_i P(A_i)$

$\Leftrightarrow P(\bigcap_{i \in K^I} A_i \bigcap \bigcap_{i \in K^I} A_i^c)$

$= \prod_{i \in K^I} P(A_i) \prod_{i \in K^I} P(A_i^c)$

$\Leftrightarrow \{\phi, \Omega, A_i, A_i^c\}$ mutual independent

$\Leftrightarrow \sigma(I\{w \in A_i\})$mutual i

$\Leftrightarrow I\{w \in A_i\}$ mutual i

2.12 X integrable $|X|$ integrable

(a) For any $A \in B(R) \Rightarrow$A is open,

so,$f^{-1}(A)$is open,so $f^{-1}(A) \in B(R)$

(b) X is known to be a random variable ,$f(x) = |x|$ continuous.

r.v. X is $\mathbf{F}/\mathbf{B}(R)$-measurable

$\Rightarrow |X|$ is $\mathbf{B}(R)/\mathbf{B}(R)$-measurable

$\Rightarrow |X|$ is $\mathbf{F}/\mathbf{B}(R)$-measurable

$\Rightarrow |X|$ is r.v.

From (a)(b),X integrable $\Leftrightarrow |X|$ integrable.

2.14

(a) Assume $\forall i, X_i$ is simple function.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_{i,j}\mathbb{I}_{A_{i,j}}\{\omega\}\right] \\
&= \int_\Omega \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_{i,j}\mathbb{I}_{A_{i,j}}\{\omega\} d\mathbb{P}(\omega) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_{i,j} \int_\Omega \mathbb{I}_{A_{i,j}}\{\omega\} d\mathbb{P}(\omega) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_{i,j}\mathbb{P}(A_{i,j}) \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i]
\end{aligned}$$

(b) Assume $\forall i, X_i$ is non-negative random variable.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sup\left\{\int_\Omega h d\mathbb{P} : h \text{ is simple and } 0 \le h \le X = \sum_{i=1}^{n} X_i\right\} \\
&= \sum_{i=1}^{n} \sup\left\{\int_\Omega h_i d\mathbb{P} : h_i \text{ is simple and } 0 \le h_i \le X_i\right\} \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i]
\end{aligned}$$

(c) Assume $\forall i, X_i$ is arbitrary random variable.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} (X_i^+ - X_i^-)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} X_i^+\right] - \mathbb{E}\left[\sum_{i=1}^{n} X_i^-\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[X_i^+\right] - \sum_{i=1}^{n} \mathbb{E}\left[X_i^-\right] \\
&= \sum_{i=1}^{n} \left(\mathbb{E}[X_i^+] - \mathbb{E}[X_i^-]\right) \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i]
\end{aligned}$$

2.15

(a) Assume X is simple function.

$$\begin{aligned}
\mathbb{E}[cX] &= \mathbb{E}\left[c \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}\{\omega\}\right] \\
&= \int_{\Omega} c \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}\{\omega\} d\mathbb{P}(\omega) \\
&= c \int_{\Omega} \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}\{\omega\} d\mathbb{P}(\omega) \\
&= c\mathbb{E}[X]
\end{aligned}$$

(b) Assume X is non-negative random variable.

$$\begin{aligned}
\mathbb{E}[cX] &= \sup\left\{\int_{\Omega} h d\mathbb{P} : h \text{ is simple and } 0 \leq h \leq cX\right\} \\
&= c \sup\left\{\int_{\Omega} h' d\mathbb{P} : h' \text{ is simple and } 0 \leq h' \leq X\right\} \\
&= c\mathbb{E}[X]
\end{aligned}$$

(c) Assume X is arbitrary random variable.

(i) $c \geq 0$

$$\begin{aligned}
\mathbb{E}[cX] &= \mathbb{E}[(cX)^+] - \mathbb{E}[(cX)^-] \\
&= \mathbb{E}[c(X)^+] - \mathbb{E}[c(X)^-] \\
&= c\mathbb{E}[(X)^+] - c\mathbb{E}[(X)^-] \\
&= c\mathbb{E}[X]
\end{aligned}$$

(ii) $c < 0$

By definition, we have

$$\begin{aligned}
(cX)^+ &= cX\mathbb{I}\{cX > 0\} \\
&= cX\mathbb{I}\{x < 0\} \text{ (since c¡0)} \\
&= (-c)(-X)\mathbb{I}\{X < 0\} \\
&= (-c)(X)^-
\end{aligned}$$

Along the similar line, we have

$$\begin{aligned}
(cX)^- &= -cX\mathbb{I}\{cX < 0\} \\
&= -cX\mathbb{I}\{X > 0\} \\
&= -c(X)^+
\end{aligned}$$

Now we can see that

$$\begin{aligned}
\mathbb{E}[cX] &= \mathbb{E}[(cX)^+] - \mathbb{E}[(cX)^-] \\
&= \mathbb{E}[(-c)(X)^-] - \mathbb{E}[-c(X)^+] \\
&= -c\mathbb{E}[(X)^-] + c\mathbb{E}[(X)^+] \\
&= c\mathbb{E}[X]
\end{aligned}$$

2.16

(a) Assume $X = \sum_{i=1}^n \alpha_i \mathbb{I}_{A_i}\{\omega\}, Y = \sum_{j=1}^m \beta_j \mathbb{I}_{B_j}\{\omega\}$ are simple functions.

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[\sum_{i=1}^n \sum_{j=1}^m \alpha_i\beta_j \mathbb{I}_{A_i}\{\omega\}\mathbb{I}_{B_j}\{\omega\}] \\
&= \int_\Omega \sum_{i=1}^n \sum_{j=1}^m \alpha_i\beta_j \mathbb{I}_{A_i}\{\omega\}\mathbb{I}_{B_j}\{\omega\} d\mathbb{P}(\omega) \\
&= \sum_{i=1}^n \sum_{j=1}^m \alpha_i\beta_j \mathbb{P}(A_i \bigcap B_j) \\
&= \sum_{i=1}^n \sum_{j=1}^m \alpha_i\beta_j \mathbb{P}(A_i)\mathbb{P}(B_j) \text{ (by the definition of independence)} \\
&= \left(\sum_{i=1}^n \alpha_i \mathbb{P}(A_i)\right)\left(\sum_{j=1}^m \beta_j \mathbb{P}(B_i)\right) \\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

(b) Assume X, Y are non-negative random variables.

$$\begin{aligned}
\mathbb{E}[XY] &= \sup\{\mathbb{E}[h] : h \text{ h is simple and } 0 \le h \le XY\} \\
&= \sup\{\mathbb{E}[h_1h_2] : h_1, h_2 \text{ are simple and } 0 \le h_1 \le X, 0 \le h_2 \le Y\} \\
&= \sup\{\mathbb{E}[h_1]\mathbb{E}[h_2] : h_1, h_2 \text{ are simple and } 0 \le h_1 \le X, 0 \le h_2 \le Y\} \\
&= \sup\{\mathbb{E}[h_1] : h_1 \text{ is simple and } 0 \le h_1 \le X\} \cdot \sup\{\mathbb{E}[h_2] : h_2 \text{ is simple and } 0 \le h_2 \le Y\} \\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

(c) Assume X, Y are arbitrary random variables.

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[(X^+ - X^-)(Y^+ - Y^-)] \\
&= \mathbb{E}[X^+Y^+ - X^+Y^- - X^-Y^+ + X^-Y^-] \\
&= \mathbb{E}[X^+]\mathbb{E}[Y^+] - \mathbb{E}[X^+]\mathbb{E}[Y^-] - \mathbb{E}[X^-]\mathbb{E}[Y^+] + \mathbb{E}[X^-]\mathbb{E}[Y^-] \\
&= (\mathbb{E}[X^+] - \mathbb{E}[X^-])(\mathbb{E}[Y^+] - \mathbb{E}[Y^-]) \\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

2.17 Before proving Ex.2.17, we need to make minor changes to the definition of conditional expectation and give a small lemma.

Definition 1. *Assume $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-algebra of $\mathcal{F}$. $X : \Omega \to \mathbb{R}$ is a random variable. The conditional expectation of X given $\mathcal{G}$ is denoted by any random variable Y which satisfies the following 2 properties:*

- *Y is $\mathcal{G}$-measurable*
- *$\forall A \in \mathcal{G}$,*

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}$$

Formally, we denoted Y by notation $\mathbb{E}[X|\mathcal{G}]$.

Lemma 1. *If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$ holds a.s.*

Proof. Since X is $\mathcal{G}$-measurable, property1 holds. And property2 holds trivially. $\square$

We can now handily prove Ex.2.17. Since $\mathbb{E}[X|\mathcal{G}_1]$ is $\mathcal{G}_1$-measurable and $\mathcal{G}_1 \subset \mathcal{G}_2$, we can see that $\mathbb{E}[X|\mathcal{G}_1]$ is $\mathcal{G}_2$-measurable. By Lemma 1, $\mathbb{E}[\mathbb{E}[X|\mathcal{G}_1]|\mathcal{G}_2] = \mathbb{E}[X|\mathcal{G}_1]$ holds almost surely.

2.18 Suppose $X = Y$ with $\mathbb{V}[X] \neq 0$. Then, we have $\mathbb{E}[XY] = \mathbb{E}[X^2] = \mathbb{V}[X] + \mathbb{E}[X]^2 \neq \mathbb{E}[X]^2 = \mathbb{E}[X]\mathbb{E}[Y]$.

2.19 As the hint suggests, $X(\omega) = \int_{[0,\infty)} \mathbb{I}\{[0, X(\omega)]\}(x)dx$. Hence, we have

$$\begin{aligned}
\mathbb{E}[X(\omega)] &= \mathbb{E}[\int_{[0,\infty)} \mathbb{I}\{[0, X(\omega)]\}(x)dx] \\
&= \int_{[0,\infty)} \mathbb{E}[\mathbb{I}\{[0, X(\omega)]\}(x)]dx \\
&= \int_{[0,\infty)} P(X(\omega) > x)dx
\end{aligned} \tag{1}$$

where the second equality is given by Fubini–Tonell theorem.

2.20 We prove the following properties all by contradiction (for the sake of rigor).

(1) Let $G = \{\omega : \mathbb{E}[X \mid \mathcal{G}](\omega) < 0\}$. Then $G \in \mathcal{G}$ since $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable by definition. Now suppose $\mathbb{P}(G) > 0$, then

$$\begin{aligned}
\int_G X d\mathbb{P} &= \int_G \mathbb{E}(X \mid \mathcal{G}) d\mathbb{P} \\
&< 0
\end{aligned} \tag{2}$$

where the equality holds by the definition of conditional expectation. Now we can find it contradictory as $X \geq 0$. Therefore $\mathbb{P}(G) = 0$, and $\mathbb{E}[X \mid \mathcal{G}] \geq 0$ a.s.

(2) Let $G = \{\omega : \mathbb{E}[1 \mid \mathcal{G}](\omega) \neq 1\}$. Then $G \in \mathcal{G}$ since $\mathbb{E}[1 \mid \mathcal{G}]$ is $\mathcal{G}$-measurable by definition. Now suppose $\mathbb{P}(G) > 0$, then

$$\begin{aligned} \int_G 1 d\mathbb{P} &= \int_G \mathbb{E}(1 \mid \mathcal{G}) d\mathbb{P} \\ &\neq 1 \end{aligned} \tag{3}$$

where the equality holds by the definition of conditional expectation. Now we can find it contradictory as $\int_G 1 d\mathbb{P} = 1$. Therefore $\mathbb{P}(G) = 0$, and $\mathbb{E}[1 \mid \mathcal{G}] = 1$ a.s.

(3) Let $G = \{\omega : \mathbb{E}[X + Y \mid \mathcal{G}](\omega) \neq \mathbb{E}[X \mid \mathcal{G}](\omega) + \mathbb{E}[Y \mid \mathcal{G}](\omega)\}$. Then $G \in \mathcal{G}$ since $\mathbb{E}[X + Y \mid \mathcal{G}]$, $\mathbb{E}[X \mid \mathcal{G}]$, and $\mathbb{E}[Y \mid \mathcal{G}]$ are all $\mathcal{G}$-measurable by definition. Now suppose $\mathbb{P}(G) > 0$, then

$$\begin{aligned} \int_G (X + Y) d\mathbb{P} &= \int_G \mathbb{E}(X + Y \mid \mathcal{G}) d\mathbb{P} \\ &\neq \int_G [\mathbb{E}(X \mid \mathcal{G}) + \mathbb{E}(Y \mid \mathcal{G})] d\mathbb{P} \\ &= \int_G \mathbb{E}(X \mid \mathcal{G}) d\mathbb{P} + \int_G \mathbb{E}(Y \mid \mathcal{G}) d\mathbb{P} \\ &= \int_G X d\mathbb{P} + \int_G Y d\mathbb{P} \end{aligned} \tag{4}$$

where the first equality and the last one hold by the definition of conditional expectation. It contradicts the linearity of expectation in that $\int_G (X + Y) d\mathbb{P} \neq \int_G X d\mathbb{P} + \int_G Y d\mathbb{P}$. Therefore $\mathbb{P}(G) = 0$, and $\mathbb{E}(X + Y \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{G}) + \mathbb{E}(Y \mid \mathcal{G})$ a.s.

(4) Let $G = \{\omega : \mathbb{E}[XY \mid \mathcal{G}](\omega) \neq Y(\omega)\mathbb{E}[X \mid \mathcal{G}](\omega)\}$. Then $G \in \mathcal{G}$ since $\mathbb{E}[XY \mid \mathcal{G}]$, Y, and $\mathbb{E}[X \mid \mathcal{G}]$ are all $\mathcal{G}$-measurable by definition. Now suppose $\mathbb{P}(G) > 0$, then

$$\begin{aligned} \int_G XY d\mathbb{P} &= \int_G \mathbb{E}(XY \mid \mathcal{G}) d\mathbb{P} \\ &\neq \int_G Y\mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} \end{aligned} \tag{5}$$

Now our target is to show it is contradictory. This is a bit tricky, so we start from the simplest case and then generalize it step by step.

a. Suppose $Y = \mathbb{I}_A$ for some $A \in \mathcal{G}$. Then

$$\int_G XY d\mathbb{P} = \int_{G \cap A} X d\mathbb{P} \tag{6}$$

and

$$\begin{aligned} \int_G Y\mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} &= \int_{G \cap A} \mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} \\ &= \int_{G \cap A} X d\mathbb{P} \end{aligned} \tag{7}$$

Hence it holds that $\int_G XY d\mathbb{P} = \int_G Y\mathbb{E}[X \mid \mathcal{G}] d\mathbb{P}$.

b. Suppose Y is non-negative and let $\{Y_n\}$ be sequence of non-negative simple functions converging to Y from below. Then by linearity, it holds that

$$\int_G X^+ Y_n d\mathbb{P} = \int_G Y_n \mathbb{E}[X^+ \mid \mathcal{G}] d\mathbb{P} \tag{8}$$

and

$$\int_G X^- Y_n d\mathbb{P} = \int_G Y_n \mathbb{E}[X^- \mid \mathcal{G}] d\mathbb{P} \tag{9}$$

Applying the monotone convergence we end up with

$$\int_G X^+ Y d\mathbb{P} = \int_G Y \mathbb{E}[X^+ \mid \mathcal{G}] d\mathbb{P} \tag{10}$$

and

$$\int_G X^- Y d\mathbb{P} = \int_G Y \mathbb{E}[X^- \mid \mathcal{G}] d\mathbb{P} \tag{11}$$

Hence,

$$\begin{aligned} \int_G XY d\mathbb{P} &= \int_G X^+ Y d\mathbb{P} - \int_G X^- Y d\mathbb{P} \\ &= \int_G Y(\mathbb{E}[X^+ \mid \mathcal{G}] - \mathbb{E}[X^- \mid \mathcal{G}]) d\mathbb{P} \\ &= \int_G Y \mathbb{E}[X^+ - X^- \mid \mathcal{G}] d\mathbb{P} \\ &= \int_G Y \mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} \end{aligned} \tag{12}$$

c. Finally, for arbitrary Y, we can separate $Y = Y^+ - Y^-$ and the contradiction still holds by linearity of expectation.

Therefore, in any case Eq.5 is contradictory. So $\mathbb{P}(G) = 0$, and $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$ a.s.

(5) Let $G = \{\omega : \mathbb{E}[X \mid \mathcal{G}_1](\omega) \neq \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1](\omega)\}$. Then $G \in \mathcal{G}_1$ since both $\mathbb{E}[X \mid \mathcal{G}_1]$ and $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1]$ are $\mathcal{G}_1$-measurable by definition. Now suppose $\mathbb{P}(G) > 0$, then

$$\begin{aligned} \int_G X d\mathbb{P} &= \int_G \mathbb{E}(X \mid \mathcal{G}_1) d\mathbb{P} \\ &\neq \int_G \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] d\mathbb{P} \\ &= \int_G \mathbb{E}(X \mid \mathcal{G}_2) d\mathbb{P} \\ &= \int_G X d\mathbb{P} \end{aligned} \tag{13}$$

The last equality stands since $G \in \mathcal{G}_1$ and $\mathcal{G}_1 \subset \mathcal{G}_2$, which suggests $G \in \mathcal{G}_2$. Now we can find it contradictory. Therefore $\mathbb{P}(G) = 0$, and $\mathbb{E}[X \mid \mathcal{G}_1] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1]$ a.s.

$$\begin{aligned} \int_G X d\mathbb{P} &= \int_G \mathbb{E}(X \mid \mathcal{G}_1) d\mathbb{P} \\ &\neq \int_G \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] d\mathbb{P} \\ &= \int_G \mathbb{E}(X \mid \mathcal{G}_2) d\mathbb{P} \\ &= \int_G X d\mathbb{P} \end{aligned} \tag{14}$$

(6) Let $G = \{\omega : \mathbb{E}\left[X \mid \sigma\left(\mathcal{G}_1 \cup \mathcal{G}_2\right)\right](\omega) \neq \mathbb{E}\left[X \mid \mathcal{G}_1\right](\omega)\}$. Notice that $\mathbb{E}\left[X \mid \mathcal{G}_1\right]$ is not only $\mathcal{G}_1$-measurable but also $\sigma\left(\mathcal{G}_1 \cup \mathcal{G}_2\right)$-measurable. Thus we have $G \in \sigma\left(\mathcal{G}_1 \cup \mathcal{G}_2\right)$. Now suppose $\mathbb{P}(G) > 0$, then

$$\begin{aligned} \int_G X d\mathbb{P} &= \int_G \mathbb{E}\left[X \mid \sigma\left(\mathcal{G}_1 \cup \mathcal{G}_2\right)\right] d\mathbb{P} \\ &\neq \int_G \mathbb{E}\left[X \mid \mathcal{G}_1\right] d\mathbb{P} \end{aligned} \tag{15}$$

To show it is contradictory, we want to prove that $\forall G \in \sigma\left(\mathcal{G}_1 \cup \mathcal{G}_2\right)$,

$$\int_G X d\mathbb{P} = \int_G \mathbb{E}\left[X \mid \mathcal{G}_1\right] d\mathbb{P} \tag{16}$$

The following techniques are closely related to 'Dynkin system', which is beyond my knowledge. The main idea is that if we assume X is non-negative, which can be generalized by linearity, it is enough to establish Eq.16 for some π-system that generates $\sigma\left(\mathcal{G}_1 \cup \mathcal{G}_2\right)$.

One possibility is $\mathcal{H} = \{G_1 \cap G_2 : G_1 \in \mathcal{G}_1, G_2 \in \mathcal{G}_2\}$. Then, $\forall G_1 \cap G_2 \in \mathcal{H}$,

$$\begin{aligned} \int_{G_1 \cap G_2} \mathbb{E}\left[X \mid \mathcal{G}_1\right] d\mathbb{P} &= \int_\Omega \mathbb{E}\left[X \mid \mathcal{G}_1\right] \mathbb{I}_{G_1} \mathbb{I}_{G_2} d\mathbb{P} \\ &= \int_\Omega \mathbb{E}\left[X \mid \mathcal{G}_1\right] \mathbb{I}_{G_1} d\mathbb{P} \int_\Omega \mathbb{I}_{G_2} d\mathbb{P} \\ &= \int_\Omega X \mathbb{I}_{G_1} d\mathbb{P} \int_\Omega \mathbb{I}_{G_2} d\mathbb{P} \\ &= \int_\Omega X \mathbb{I}_{G_1} \mathbb{I}_{G_2} d\mathbb{P} \\ &= \int_{G_1 \cap G_2} X d\mathbb{P} \end{aligned} \tag{17}$$

where the second and fourth equality holds due to independence between $\sigma(X)$ and $\mathcal{G}_2$ given $\mathcal{G}_1$.

Hence, we find it contradictory. So $\mathbb{P}(G) = 0$ and $\mathbb{E}\left[X \mid \sigma\left(\mathcal{G}_1 \cup \mathcal{G}_2\right)\right] = \mathbb{E}\left[X \mid \mathcal{G}_1\right]$ a.s.

(7) Let $G = \{\omega : \mathbb{E}[X \mid \mathcal{G}](\omega) \neq \mathbb{E}[X]\}$. Then $G \in \mathcal{G}$ since $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable by definition. And because $\mathcal{G}$ is trivial, $G = \emptyset$ or $G = \Omega$.

a. If $G = \emptyset$, $P(G) = 0$ for sure.

b. If $G = \Omega$, which suggests $\mathbb{E}[X \mid \mathcal{G}] \neq \mathbb{E}[X]$ always holds, we have

$$\begin{aligned} \int_G X d\mathbb{P} &= \int_G \mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} \\ &\neq \int_G \mathbb{E}[X] d\mathbb{P} \\ &= \int_\Omega \mathbb{E}[X] d\mathbb{P} \\ &= \mathbb{E}[X] \end{aligned} \tag{18}$$

which is obviously contradictory since $\int_G X d\mathbb{P} = \int_\Omega X d\mathbb{P} = \mathbb{E}[X]$.

Therefore, $P(G) = 0$ and hence $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$ a.s.

Chapter 3 Stochastic Processes and Markov Chains

3.1

(a) On$([0,1],\mathcal{B},\lambda)$,for any $x \in [0,1]$

Let $F_1(x), F_2(x), F_3(x)$,...be the binary expansion of x.

$F_t(x)$= $\begin{cases} 1, A \\ 0, \overline{A} \quad (\overline{A} \text{ is the opposite case of } A) \end{cases}$

$F_t(x)$ is Bernoulli random variable.

(b) $\begin{cases} F_1 = 0 : 0 \leq x < 0.5 \\ F_1 = 1 : 0.5 \leq x < 1 \end{cases}$

$\begin{cases} F_2 = 0 : 0 \leq x' < 0.5 \\ F_2 = 1 : 0.5 \leq x' < 1 \end{cases}$ x'=2x-1

...

$\begin{cases} F_t = 0 : 0 \leq x^t < 0.5 \Rightarrow \mathbb{P}(F_t = 0) = \frac{1}{2} \\ F_t = 1 : 0.5 \leq x^t < 1 \Rightarrow \mathbb{P}(F_t = 1) = \frac{1}{2} \end{cases}$

(c) It is obviously that $(F_t)_{t=1}^{\infty}$ are independent. It satisfies independent equation: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

(d) $(X_{m,t})_{t=1}^{\infty}$ is a subsequence of $(F_t)_{t=1}^{\infty}$ and $(X_{m,t})_{t=1}^{\infty}$ are mutually exclusive.

(e) Such as(d).

(f) Such as(d).

3.2

(a) $S_t = \sum_{s=1}^{t} X_s 2^{s-1}$

X_t is a F-adapted martingale.

(1)$\mathbb{E}[X_t|\mathcal{F}_{t-1}] = X_{t-1}$.

(2)X_t is integrable $\Rightarrow$ S_t is integrable.

$$\begin{aligned}
\mathbb{E}[S_t|\mathcal{F}_{t-1}] &= \mathbb{E}[S_{t-1} + X_t 2^{t-1}|\mathcal{F}_{t-1}] \\
&= S_{t-1} + \mathbb{E}[X_t 2^{t-1}|\mathcal{F}_{t-1}] \\
&= S_{t-1} + 2^t \times (1) \times \frac{1}{2} + 2^t \times (-1) \times \frac{1}{2} \\
&= S_{t-1}
\end{aligned}$$

$\Rightarrow (S_t)_{t=1}^{\infty}$

(b) t=1 , if $S_t \neq 1 \Rightarrow X_1 = -1, S_t$=-1
t=2 , if $S_t \neq 1 \Rightarrow X_1 = -1, S_t$=-3
t=3 , if $S_t \neq 1 \Rightarrow X_1 = -1, S_t$=-7
...
If avoid S_t=1 , the X_s sequence must be -1.
$\tau = \min\{t : S_t = 1\} = \min\{t : X_T = 1\}$
$\Rightarrow \mathbb{P}(\tau < n) = 1 - \mathbb{P}(\tau \geq n) = 1 - \frac{1}{2^n}$
$\Rightarrow \mathbb{P}(\tau < \infty) = 1 - \lim_{n\to\infty} \mathbb{P}(\tau \geq n) = 1 - \frac{1}{2^n} = 1 - \lim_{n\to\infty} \frac{1}{2^n}$

(c) If t=τ , then S_t=1 , so $S_\tau \equiv 1$
$\Rightarrow \mathbb{E}[S_\tau] = 1$

(d) Doob's(a)can be proved by 3.2(b)
$\tau = 1 \Rightarrow X_1 = 1 \Rightarrow \mathbb{P}(\tau = 1) = \frac{1}{2}$
$\tau = 2 \Rightarrow X_1 = -1X_2 = 1 \Rightarrow \mathbb{P}(\tau = 1) = \frac{1}{4}$
$\tau = 3 \Rightarrow X_1 = -1X_2 = -1X_3 = 1 \Rightarrow \mathbb{P}(\tau = 1) = \frac{1}{8}$
...
$\mathbb{P}(\tau < \infty) = \mathbb{P}(\tau = 1) + \mathbb{P}(\tau = 2) + \mathbb{P}(\tau = 3) + ... = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + ... = 1$
because of $n \neq \infty, \mathbb{P}(\tau = n) = \frac{1}{n^2} \neq 0$.
Doob's(b)(c)can also be proved by 3.2(b)
t=1 , if $S_t \neq 1 \Rightarrow X_1 = -1, S_t = -1$
t=2 , if $S_t \neq 1 \Rightarrow X_1 = -1, S_t = -3$
t=3 , if $S_t \neq 1 \Rightarrow X_1 = -1, S_t = -7$
...
It can be concluded that —S_t— and —S_{t-1}— can not be bounded, so $\mathbb{E}[|X_{t+1}|\mathcal{F}]$ and $|4X_{t\wedge\tau}|$ can not be bounded neither.

3.4 If $X_t \geq 0$ is dropped, $\mathbb{E}[X_\tau|\{\tau \leq n\}] \geq \mathbb{E}[\varepsilon|\{\tau \leq n\}]$ not always true.

Chapter 4 Stochastic Bandits

4.1 By definition

$$\begin{aligned}
R_n(\pi, v) &= n\mu^*(v) - \mathbb{E}[\sum_{t=1}^{n} X_t] \\
&= \sum_{t=1}^{n} \mu^*(v) - \sum_{t=1}^{n} \mathbb{E}[X_t] \\
&= \sum_{t=1}^{n} [\mu^* - \mu_{A_t}]
\end{aligned}$$

(a) $\mu^* = \max \mu_a \geq \mu_{A_t} \Rightarrow R_n(\pi, v) = \sum_{t=1}^{n} [\mu^* - \mu_{A_t}] \geq 0$.

(b) If π choose $A_t \in \arg\max_a \mu_a$ for all $t \in [n] \Rightarrow \sum_{t=1}^{n} [\mu^* - \mu_{A_t}] = 0$.

(c) If $R_n(\pi, v) = 0$ for some policy π ,then $A_t \in \arg\max_a \mu_a \Rightarrow \mathbb{P}(\mu_{A_t} = \mu^*) = 1$.

4.3 Denote $h_t = a_1, x_1, \ldots, a_t, x_t$.

(a) According to the definition of conditional probability and marginal distribution, we have

$$\begin{aligned}
p_{v\pi}(a_n \mid h_{n-1}) &= \frac{p_{v\pi}(h_{n-1}, a_n)}{p_{v\pi}(h_{n-1})} \\
&= \frac{\int_{\mathbb{R}} p_{v\pi}(h_n) dx_n}{p_{v\pi}(h_{n-1})} \\
&= \frac{\int_{\mathbb{R}} \prod_{t=1}^{n} \pi\left(a_t \mid h_{t-1}\right) p_{a_t}\left(x_t\right) dx_n}{p_{v\pi}(h_{n-1})} \\
&= \frac{\prod_{t=1}^{n-1} \pi\left(a_t \mid h_{t-1}\right) p_{a_t}\left(x_t\right)}{p_{v\pi}(h_{n-1})} \int_{\mathbb{R}} \pi\left(a_n \mid h_{n-1}\right) p_{a_n}\left(x_n\right) dx_n \\
&= \pi\left(a_n \mid h_{n-1}\right) \int_{\mathbb{R}} p_{a_n}\left(x_n\right) dx_n \\
&= \pi\left(a_n \mid h_{n-1}\right)
\end{aligned}$$

(b) According to the definition of conditional probability and marginal distribution, we have

$$\begin{aligned} p_{v\pi}(x_n \mid h_{n-1}, a_n) &= \frac{p_{v\pi}(h_n)}{p_{v\pi}(h_{n-1}, a_n)} \\ &= \frac{p_{v\pi}(h_n)}{\int_{\mathbb{R}} p_{v\pi}(h_n) dx_n} \\ &= \frac{p_{v\pi}(h_n)}{\int_{\mathbb{R}} \left[\prod_{t=1}^{n} \pi\left(a_t \mid h_{t-1}\right) p_{a_t}\left(x_t\right)\right] dx_n} \\ &= \frac{p_{v\pi}(h_n)}{\prod_{t=1}^{n-1} \pi\left(a_t \mid h_{t-1}\right) p_{a_t}\left(x_t\right)} \frac{1}{\int_{\mathbb{R}} \pi\left(a_n \mid h_{n-1}\right) p_{a_n}\left(x_n\right) dx_n} \\ &= \pi\left(a_n \mid h_{n-1}\right) p_{a_n}\left(x_n\right) \frac{1}{\pi\left(a_n \mid h_{n-1}\right)} \\ &= p_{a_n}\left(x_n\right) \end{aligned}$$

4.4 Denote $h_t = a_1, x_1, \ldots, a_t, x_t$. The policy that mixes the policies can be defined as

$$\pi_t^{\circ}\left(a_t \mid h_{t-1}\right) = \frac{\sum_{\pi \in \Pi} p(\pi) \prod_{s=1}^{t} \pi_s\left(a_s \mid h_{s-1}\right)}{\sum_{\pi \in \Pi} p(\pi) \prod_{s=1}^{t-1} \pi_s\left(a_s \mid h_{s-1}\right)}$$

.

By the definition of the canonical probability space and the product of probability kernels,

$$\begin{aligned} \mathbb{P}_{v\pi^{\circ}}(B) &= \sum_{a_1=1}^{k} \int_{\mathbb{R}} \cdots \sum_{a_n=1}^{k} \int_{\mathbb{R}} \mathbb{I}_B\left(h_n\right) v_{a_n}\left(dx_n\right) \pi_n^{\circ}\left(a_n \mid h_{n-1}\right) \cdots v_{a_1}\left(dx_1\right) \pi_1^{\circ}\left(a_1\right) \\ &= \sum_{\pi \in \Pi} p(\pi) \sum_{a_1=1}^{k} \int_{\mathbb{R}} \cdots \sum_{a_n=1}^{k} \int_{\mathbb{R}} \mathbb{I}_B\left(h_n\right) v_{a_n}\left(dx_n\right) \pi_n\left(a_n \mid h_{n-1}\right) \cdots v_{a_1}\left(dx_1\right) \pi_1\left(a_1\right) \\ &= \sum_{\pi \in \Pi} p(\pi) \mathbb{P}_{v\pi}(B), \end{aligned}$$

where the second equality follows by substituting the definition of π_n° and induction.

Chapter 5 Concentration of Measure

5.1

$$V(\hat{\mu}) = E((\hat{\mu}-\mu)^2) = E((\frac{1}{n}\sum_{t=1}^{n} X_t - \mu)^2) = E(\frac{1}{n^2}\sum_{t=1}^{n}(X_t-\mu)^2) = \frac{1}{n^2}\sum_{t=1}^{n} E(X_t-\mu)^2 = \frac{1}{n^2}\sum_{t=1}^{n}\sigma^2 = \frac{\sigma^2}{n} \tag{19}$$

5.4

(a)

$$P(|X| \geq \varepsilon) = P(X \geq \varepsilon)I\{X \geq 0\} + P(X \leq -\varepsilon)I\{X < 0\} = \int_{\varepsilon}^{\infty} \frac{x}{2} exp\{\frac{-x^2}{2}\}dx + \int_{-\infty}^{\varepsilon} \frac{-x}{2} exp\{\frac{-x^2}{2}\}dx \tag{20}$$

Calculate the above formula and get the result ,
$P(|X| \geq \varepsilon) = \frac{1}{2}exp\{\frac{-\varepsilon^2}{2}\} + \frac{1}{2}exp\{\frac{-\varepsilon^2}{2}\}$
$= exp\{\frac{-\varepsilon^2}{2}\}$
(b)
Let's start with a lemma:
If X is $\sigma-$subgaussian,then $P(|X| > t) \leq exp\{-b\varepsilon^2\}$, where $b = exp\{-\sigma^2\}$
The proof of lemma is omitted.
It can be seen from the first question , $P(|X| \geq \varepsilon) = exp\{\frac{-\varepsilon^2}{2}\}$
The comparison of the two formulas shows that , $0 < b \leq 1/2$. That is, $\sigma \geq \sqrt{ln2}$
By topic condition , $\sigma = \sqrt{2-\varepsilon}$
Hence , $\varepsilon \leq 2 - ln2$, this is in contradiction with the arbitrariness of ε

5.7

(a)If X is $\sigma-$subgaussian , then $E(X) = 0$,$E(X^2) \leq \sigma^2$
proof:

$$E(e^{\lambda X}) = \sum_{n=0}^{\infty} \frac{\lambda^n E(X^n)}{n!} = 1 + \lambda E(X) + \frac{\lambda^2 E(X^2)}{2} + O(\lambda^2) \tag{21}$$

By definition ,

$$E(e^{\lambda X}) \leq e^{\frac{\lambda^2\sigma^2}{2}} = 1 + \frac{\lambda^2\sigma^2}{2} + O(\lambda^2) \tag{22}$$

By comparing the above two formulas and discussing the case that a approaches to 0 from above and below 0, we get the conclusion that ,
$E(X) = 0$,$E(X^2) \leq \sigma^2$
(b)

If X is $\sigma-$subgaussian , then $E(X)=0$,$E(X^2)\leq\sigma^2$.
$E(e^{c\lambda x})=1+\lambda E(cx)+\frac{\lambda^2E(c^2x^2)}{2}+O(\lambda^2)$
$\leq 1+c\lambda E(x)+\frac{\lambda^2c^2}{2}E(x^2)+O(\lambda^2)$
$\leq 1+\frac{\lambda^2c^2\sigma^2}{2}+O(\lambda^2)$
$\leq e^{\frac{\lambda^2c^2\sigma^2}{2}}$
Hence , cX is $|c|\sigma-$subgaussian .
(c)
If X_1 is σ_1-subgaussian , X_2 is σ_2-subgaussian
then $E(X_1)=0$,$E(X_1^2)\leq\sigma_1^2$,$E(X_2)=0$,$E(X_2^2)\leq\sigma_2^2$
$E(e^{\lambda(x_1+x_2)})=1+\lambda E(x_1+x_2)+\frac{\lambda^2E((x_1+x_2)^2)}{2}+O(\lambda^2)$
$=1+\frac{\lambda^2}{2}Var(x_1+x_2)+O(\lambda^2)$
$=1+\frac{\lambda^2}{2}(var(x_1)+var(x_2)+2cov(x_1,x_2))+O(\lambda^2)$
Because x_1, x_2 are independent ,
$=1+\frac{\lambda^2}{2}(E(x_1^2)+E(x_2^2))(\lambda^2)$
$\leq 1+\frac{\lambda^2}{2}(\sigma_1^2+\sigma_2^2)+O(\lambda^2)$
$\leq e^{\frac{\lambda^2(\sigma_1^2+\sigma_2^2)}{2}}$
Hence , X_1+X_2 is $\sqrt{\sigma_1^2+\sigma_2^2}-$subgaussian .

5.11
(a)

$$E(e^{\lambda X})=1+\lambda E(X)+\frac{\lambda^2E(X^2)}{2}+O(\lambda^2)=1+\frac{\lambda^2E(X^2)}{2}+O(\lambda^2) \tag{23}$$

If the conclusion is true, then the above formula satisfies
$\leq 1+\frac{\lambda^2}{2}(\frac{(b-a)^2}{4})+O(\lambda^2)$
So just prove:
$E(x^2)\leq(\frac{b-a}{2})^2$
$E(x^2)=var(x)=E(x-\bar{x})^2$
However,$(x-\bar{x})^2\leq(\frac{b-a}{2})^2$. The conclusion is proved.
(b)
The proof of Hoeffding's Inequality:
Let $X_i=Z_i-E(Z_i)$, $\bar{X}=\frac{1}{m}\sum_{i=1}^m X_i$
By Markov inequality , for all $\lambda>0$, $\varepsilon>0$,
$P(\bar{X}\geq\varepsilon)=P(e^{\lambda\bar{X}}\geq e^{\lambda\varepsilon})\leq\frac{E(e^{\lambda\bar{X}})}{e^{\lambda\varepsilon}}$
Z_1,$\cdots$,Z_m iid.r.v.
So,$E(e^{\lambda\bar{X}})=\prod_{i=1}^m E(e^{\frac{\lambda X_i}{m}})$
By Hoeffding's lamma,
$E(e^{\frac{\lambda X_i}{m}})\leq e^{\frac{\lambda^2(b-a)^2}{8m^2}}$
So , $P(\bar{X}\geq\varepsilon)\leq e^{-\lambda\varepsilon}\prod_{i=1}^m E(e^{\frac{\lambda X_i}{m}})$
$\leq e^{-\lambda\varepsilon}e^{\frac{\lambda^2(b-a)^2}{8m}}$
$\leq e^{-\lambda\varepsilon+\frac{\lambda^2(b-a)^2}{8m}}$
Let $\lambda=\frac{4m\varepsilon}{(b-a)^2}$, then $P(\bar{X}\geq\varepsilon)\leq e^{\frac{-2m\varepsilon^2}{(b-a)^2}}$
Similarly, we can prove the other side of the inequality.

5.16 By assumption $Pr(X_t\leq x)\leq x$, which means that for$\lambda<1$,

$$\mathbb{E}\left[exp(\lambda log(\frac{1}{x_t}))\right]=\int_0^\infty P(exp(\lambda log(\frac{1}{x_t}))\geq x)dx=1+\int_1^\infty P(X_t\leq x^{-\frac{1}{\lambda}})dx \tag{24}$$

Applying the Cramer-Chernoff method,

$$P\left(\sum_{t=1}^n log(\frac{1}{X_t})\geq\epsilon\right)=P\left(exp(\lambda\sum_{t=1}^n log(\frac{1}{X_t}))\geq exp(\lambda\epsilon)\right)\leq\left(\frac{1}{1-\lambda}\right)^n exp(-\lambda\epsilon)$$

choosing $\lambda = \frac{\epsilon - n}{\epsilon}$ completes the claim.
5.18(a) Let $\lambda > 0$. Then,

$$exp(\lambda \mathbb{E}[Z]) \leq \mathbb{E}[exp(\lambda Z)] \leq \sum_{t=1}^{n} \mathbb{E}[exp(\lambda X_t)] \leq nexp(\frac{\lambda^2 \sigma^2}{2})$$

Rearranging shows that,

$$\mathbb{E}(Z) \leq \frac{log(n)}{\lambda} + \frac{\lambda \sigma^2}{2}$$

Choosing $\lambda = \frac{1}{\sigma}\sqrt{2log(n)}$ shows that $\mathbb{E}(Z) \leq \sqrt{2\sigma^2 log(n)}$

Chapter 6 The Explore-Then-Commit Algorithm

6.2 We proceed by comparing the values of $n\Delta$ and $\Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{n\Delta^2}{4}\right)\right\}\right)$.

(a) If $n\Delta > \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{n\Delta^2}{4}\right)\right\}\right)$, we have $(n-1)\Delta^2 > 4(1 + \max\left\{0, \log\left(\frac{n\Delta^2}{4}\right)\right\}) \geq 4$, which suggests that $\Delta \geq \frac{2}{\sqrt{n}}$. Therefore,

$$\begin{aligned}
R_n &= \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{n\Delta^2}{4}\right)\right\}\right) \\
&= \Delta + \frac{4}{\Delta}\left(1 + \log\left(\frac{n\Delta^2}{4}\right)\right) \\
&= \Delta + \frac{4}{\Delta} + \frac{4}{\Delta}\log(\frac{n\Delta^2}{4}) \\
&\leq \Delta + 2\sqrt{n} + \frac{16}{e^4}\sqrt{n} \\
&= \Delta + C\sqrt{n},
\end{aligned}$$

where the inequality follows from taking $x^* = \frac{2e^4}{\sqrt{n}}$ to maximize $f(x) = \frac{4}{x}log(\frac{nx^2}{4})$.

(a) If $n\Delta \leq \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{n\Delta^2}{4}\right)\right\}\right)$, we consider another two cases:

(i) If $\Delta \geq \frac{2}{\sqrt{n}}$, by (1) we still have $R_n = n\Delta \leq \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{n\Delta^2}{4}\right)\right\}\right) \leq \Delta + C\sqrt{n}$.

(ii) If $\Delta < \frac{2}{\sqrt{n}}$, $R_n \leq n\Delta \leq 2\sqrt{n} \leq \Delta + C\sqrt{n}$, where the first inequality is trivial.

6.3 Suppose $\Delta_1 = 0$, $\Delta_2 = \Delta > 0$. Then, the probability that we choose the suboptimal arm (i.e., the second arm) after commitment is

$$\begin{aligned}
\mathbb{P}(T_2(n) > m) &= \mathbb{P}(\hat{\mu}_2(2m) > \hat{\mu}_1(2m)) \\
&= \mathbb{P}([\hat{\mu}_2(2m) - \mu_2] - [\hat{\mu}_1(2m) - \mu_1] > \Delta \\
&\leq \exp(-\frac{m\Delta^2}{4}),
\end{aligned}$$

where the inequality follows from Theorem 5.3. By letting $\exp(-\frac{m\Delta^2}{4}) = \delta$, we have $m = -\frac{4\log\delta}{\Delta^2}$. Hence, if we take $m = \min\{\lfloor\frac{n}{2}\rfloor, -\frac{4\log\delta}{\Delta^2}\}$, with high probability we have

$$\begin{aligned}
\bar{R}_n &= \Delta T_2(n) \\
&\leq \Delta m \\
&= \min\{\lfloor\frac{n}{2}\rfloor\Delta, -\frac{4\log\delta}{\Delta}\}
\end{aligned}$$

6.4 Denote the reward received in the t-th interaction with arm i as $X_{i,t}$. From 6.3 we have that with probability $1-\delta$,

$$\begin{aligned}\hat{R}_n &\leq \sum_{t=1}^{n-m}(\mu_1 - X_{1,t}) + \sum_{t=1}^{m}(\mu_1 - X_{2,t}) \\ &= \sum_{t=1}^{n-m}(\mu_1 - X_{1,t}) + \sum_{t=1}^{m}(\mu_2 - X_{2,t}) + m\Delta\end{aligned}$$

Notice that the sum of the first two terms is $(\sqrt{(n-m)^2+m^2})$-subgaussian. Therefore, with probability $(1-\delta)^2$, we have $\hat{R}_n \leq \sqrt{-2[(n-m)^2+m^2]\log\delta} + m\Delta$. This suggests that compared to that derived for the pseudo-regret, the bound on the random regret is less tight with a smaller probability as more randomness is considered.

6.5 Suppose $\Delta_1 = 0$, $\Delta_2 = \Delta > 0$.

(a) By Theorem 6.1, we have

$$\begin{aligned}R_n(v) &= \Delta\mathcal{E}[T_2(n)] \\ &\leq m\Delta + (n-2m)\Delta\exp(-\frac{m\Delta^2}{4}) \\ &\leq m\Delta + n\Delta\exp(-\frac{m\Delta^2}{4}) \\ &\leq m\Delta + n\sqrt{\frac{2}{m}}\exp(-\frac{1}{2}) \\ &= [\Delta + \sqrt{2}\exp(-\frac{1}{2})]n^{\frac{2}{3}},\end{aligned}$$

where the last inequality follows from taking $x^* = \sqrt{\frac{2}{m}}$ to maximize $f(x) = x\exp(-\frac{m\Delta^2}{4})$, and the last equality follows from taking $m = n^{\frac{2}{3}}$.

Assume there is such a $C > 0$ that leads to $R_n(v) \leq \Delta_v + Cn^{2/3}$ for any problem instance v and $n \geq 1$. Since trivially $R_n(v) \geq m\Delta$, we have $m\Delta \leq \Delta + Cn^{\frac{2}{3}} \Rightarrow m \leq 1 + \frac{Cn^{\frac{2}{3}}}{\Delta}$. Under this circumstance, we can easily find a problem instance v with $\Delta \to \infty$ such that $m \leq 1$. Recalling we only explore $2m$ rounds, we will eventually pull the suboptimal arm with a high probability, which contradicts our assumption.

(c) We proceed by comparing the values of $\frac{C\log n}{\Delta}$ and $C\sqrt{n\log(n)}$.

(i) If $\Delta \geq \sqrt{\frac{\log n}{n}}$, $R_n(v) \leq \Delta + C\frac{\log n}{\Delta} \leq \Delta + C\sqrt{n\log n}$.

(ii) If $\Delta < \sqrt{\frac{\log n}{n}}$, $R_n(v) \leq n\Delta \leq \sqrt{n\log n} \leq \Delta + C\sqrt{n\log n}$.

(e) We proceed by comparing the values of e and $n\Delta^2$.

(i) If $\Delta \geq \sqrt{\frac{e}{n}}$, $R_n(v) \leq \Delta + \frac{C\log(n\Delta^2)}{\Delta} \leq \Delta + \frac{2C}{e}\sqrt{n} = \Delta + C\sqrt{n}$.

(ii) If $\Delta < \sqrt{\frac{e}{n}}$, $R_n(v) \leq n\Delta \leq \sqrt{en} \leq \Delta + C\sqrt{n}$.

Chapter 11 The Exp3 Algorithm

11.2 Let π be a deterministic policy, and we define $x_{ti} = 0$ if $A_t = i$ otherwise $x_{ti} = 1$. The deterministic policy collects zero rewards all time,

$$\max_{i \in [k]} \sum_{t=1}^{n} x_{ti} \geq \frac{1}{k} \sum_{t=1}^{n} \sum_{i=1}^{k} x_{ti} = \frac{n(k-1)}{k}$$

11.5 Let P be a probability vector with nonzero components and let $A \sim P$. Suppose $\hat{X}$ is a function such that for all $x \in \mathbb{R}^k$,

$$\mathbb{E}\left[\hat{X}\left(A, x_A\right)\right] = \sum_{i=1}^{k} P_i \hat{X}\left(i, x_i\right) = x_1$$

Show that there exists an $a \in \mathbb{R}^k$ such that $< a, P >= 0$ and for all i and z in their respective domains, $\hat{X}(i,z) = a_i + \frac{\mathbb{I}\{i=1\}z}{P_1}$

Proof. Let x, x' be arbitrary but agree on the first component $x_1 = x'_1$. Let $f(x) = \sum_{i=1}^{k} P_i \hat{X}\left(i, x_i\right)$ Note that,

$$0 = f(x) - f(x') = \sum_{i=j}^{k} P_j \hat{X}\left(j, x_j\right)$$

for all $j > 1$. Since x, x' are arbitrary, $\hat{X}(j,) = const$. Let a_j equal to $\hat{X}(j,)$.

Further, let $a_1 = \hat{X}(1,0)$ and then given any $x_1 \in \mathbb{R}$, $\hat{X}\left(1, x_1\right) = a_1 + x_1/P_1$.

Finally, let x be such that $x_1 = 0$. Then $0 = f(x) = \sum_i P_i a_i$. □

11.7 First, note that if $G = -\log(-\log(U))$ then $\mathbb{P}(G \leq g) = e^{-\exp(-g)}$.

$$\begin{aligned}
\mathbb{P}\left(\log a_i + G_i \geq \max_{j \in [k]} \log a_j + G_j\right) &= \mathbb{E}\left[\prod_{j \neq i} \mathbb{P}\left(\log a_j + G_j \leq \log a_i + G_i \mid G_i\right)\right] \\
&= \mathbb{E}\left[\prod_{j \neq i} \exp\left(-\frac{a_j}{a_i} \exp\left(-G_i\right)\right)\right] \\
&= \mathbb{E}\left[U_i^{\sum_{j \neq i} \frac{a_j}{a_i}}\right] \\
&= \frac{1}{1 + \sum_{j \neq i} \frac{a_j}{a_i}} \\
&= \frac{a_i}{\sum_{j=1}^{k} a_j}
\end{aligned}$$

11.8 Let $Z_t i$ be a standard Gambel. The follow-theperturbed-leader algorithm chooses

$$A_t = \text{argmax}_{i \in [k]} \left(Z_{ti} - \eta \sum_{s=1}^{t-1} \hat{Y}_{si}\right)$$

is the same as EXP3. Given (11.7)

$$\mathbb{P}\left(\log\left(a_i\right) + G_i = \max_{j\in[k]}\left(\log\left(a_j\right) + G_j\right)\right) = \frac{a_i}{\sum_{j=1}^{k} a_j}$$

Just simply take a_i as $-\eta \sum_{s=1}^{t-1} \hat{Y}_{si}$, then the form is identical.

Chapter 18

18.1

(a) By Jensen's inequality,

$$\begin{aligned}
\sum_{c\in\mathcal{C}} \sqrt{\sum_{t=1}^{n} \mathcal{I}\{c_t = c\}} &= \|C\| \sum_{c\in\mathcal{C}} \frac{1}{\|C\|} \sqrt{\sum_{t=1}^{n} \mathcal{I}\{c_t = c\}} \\
&\leq \|C\| \sqrt{\sum_{c\in\mathcal{C}} \frac{1}{\|C\|} \sum_{t=1}^{n} \mathcal{I}\{c_t = c\}} \\
&= \sqrt{\|C\| n}
\end{aligned}$$

(b) When each context occurs $\frac{n}{\|C\|}$ times we have

$$\sum_{c\in\mathcal{C}} \sqrt{\sum_{t=1}^{n} \mathcal{I}\{c_t = c\}} = \sqrt{n\|C\|}$$