Part III

Appendix

B Measure Theory

B.1 Measure and Probability

The foundational concept in measure theory is the sigma-algebra, which defines the collection of subsets to which we can assign a measure.

Definition B.1. A σ-algebra $\mathcal{F}$ on a set Ω is a collection of subsets of Ω that satisfies the following properties:

(i) $\varnothing \in \mathcal{F}$

(ii) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$

(iii) If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Once we have a σ-algebra, we can define a measure, which generalizes the concepts of length, area, and probability.

Definition B.2. Consider a measurable space $(\Omega, \mathcal{F})$. A *measure* μ on a σ-algebra $\mathcal{F}$ is a function that assigns a non-negative real number to each set in $\mathcal{F}$ and satisfies the following properties:

1. $\mu(\varnothing) = 0$
2. If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

If $\mu(\Omega) < \infty$, then μ is called a *finite measure*. If in addition $\mu(\Omega) = 1$, then μ is a *probability measure*.

These components form the standard objects of study in measure theory.

Definition B.3. A pair $(\Omega, \mathcal{F})$ consisting of a set Ω and a σ-algebra $\mathcal{F}$ is called a *measurable space*. A triple $(\Omega, \mathcal{F}, \mu)$ consisting of a measurable space and a measure μ is called a *measure space*. If μ is a probability measure, the triple is called a *probability space*.

Many important measures are not finite, but satisfy a weaker condition called σ-finiteness.

Definition B.4. A measure μ on $(\Omega, \mathcal{F})$ is called *σ-finite* if there exists a sequence of sets $A_1, A_2, \ldots \in \mathcal{F}$ such that $\bigcup_{i=1}^{\infty} A_i = \Omega$ and $\mu(A_i) < \infty$ for all i.

A simple example of a measure that can be finite or σ-finite is the counting measure.

Example B.5 (Counting Measure)**.** Let Ω be a countable set and $\mathcal{F} = 2^{\Omega}$. The counting measure μ is defined by $\mu(A) = |A|$ (the number of elements in A) for any $A \subseteq \Omega$. This measure is σ-finite since Ω is countable (take $A_i = \{\omega_i\}$). ♢

Measures are often defined on a smaller class of sets (like intervals in $\mathbb{R}$) and then extended to the full σ-algebra. Carathéodory's Extension Theorem guarantees that this extension is unique for σ-finite measures.

Theorem B.6 (Uniqueness of Measure Extension)**.** *Let $\mathcal{A}$ be a collection of subsets of Ω that is closed under finite intersections (a π-system) and generates the σ-algebra $\mathcal{F} = \sigma(\mathcal{A})$. If two measures μ and ν on $(\Omega, \mathcal{F})$ agree on $\mathcal{A}$ (i.e., $\mu(A) = \nu(A)$ for all $A \in \mathcal{A}$), and they are σ-finite on $\mathcal{A}$, then $\mu = \nu$ on $\mathcal{F}$.*

Measures also behave continuously with respect to increasing or decreasing sequences of sets.

Proposition B.7. *Let μ be a measure on $(\Omega, \mathcal{F})$.*

1. ***(Continuity from below)*** *If $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing sequence of sets in $\mathcal{F}$ and $A = \bigcup_{n=1}^{\infty} A_n$, then*

$$\mu(A) = \lim_{n\to\infty} \mu(A_n).$$

2. ***(Continuity from above)*** *If $A_1 \supseteq A_2 \supseteq \cdots$ is a decreasing sequence of sets in $\mathcal{F}$ with $\mu(A_1) < \infty$ and $A = \bigcap_{n=1}^{\infty} A_n$, then*

$$\mu(A) = \lim_{n\to\infty} \mu(A_n).$$

We now turn to the functions between measurable spaces, which must preserve the measurable structure.

Definition B.8. Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{G})$ be measurable spaces. A function $f : \Omega \to S$ is *measurable* (or $\mathcal{F}/\mathcal{G}$-measurable) if for every $B \in \mathcal{G}$, the preimage $f^{-1}(B) \in \mathcal{F}$.

That is, f is measurable if

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F} \quad \text{for all } B \in \mathcal{G}.$$

Conversely, any function induces a σ-algebra on its domain.

Definition B.9. Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{G})$ be measurable spaces, and let $f : \Omega \to S$ be a measurable function. The *σ-algebra generated by f*, denoted by $\sigma(f)$, is the collection of all preimages of sets in $\mathcal{G}$:

$$\sigma(f) = \{f^{-1}(B) : B \in \mathcal{G}\}.$$

This is the smallest σ-algebra on Ω with respect to which f is measurable. Note that $\sigma(f) \subseteq \mathcal{F}$ since f is measurable.

Definition B.10. The *Borel σ-algebra* on a topological space $(X, \mathcal{T})$, denoted by $\mathcal{B}(X)$, is the σ-algebra generated by the open sets $\mathcal{T}$. In particular, if (X, d) is a metric space, $\mathcal{B}(X)$ is generated by the open balls.

For $X = \mathbb{R}$, $\mathcal{B}(\mathbb{R})$ is the σ-algebra generated by the collection of all open intervals in $\mathbb{R}$. Sets in $\mathcal{B}(\mathbb{R})$ are called *Borel sets*. This is the standard σ-algebra used when the sample space is $\mathbb{R}$ (or $\mathbb{R}^d$).

On the real line, the most important measure is the one that assigns lengths to intervals.

Definition B.11 (Lebesgue Measure)**.** The Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is the unique measure satisfying $\lambda((a, b]) = b - a$ for all intervals $(a, b]$.

The Lebesgue measure is σ-finite since $\mathbb{R} = \bigcup_{n=1}^{\infty}(-n, n]$.

Measurable functions are closed under various operations.

Proposition B.12. *Let $(\Omega, \mathcal{F})$, $(S, \mathcal{G})$, and $(T, \mathcal{H})$ be measurable spaces.*

1. *(**Composition**) If $f : \Omega \to S$ is $\mathcal{F}/\mathcal{G}$-measurable and $g : S \to T$ is $\mathcal{G}/\mathcal{H}$-measurable, then the composition $g \circ f : \Omega \to T$ is $\mathcal{F}/\mathcal{H}$-measurable.*

A measurable function can be used to transport a measure from its domain to its codomain.

Definition B.13 (Push-forward Measure)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $(S, \mathcal{G})$ a measurable space, and $T : \Omega \to S$ a measurable function. The *push-forward measure* of μ by T, denoted μ^T (or sometimes $T_{\#}\mu$ or $\mu \circ T^{-1}$), is the measure on $(S, \mathcal{G})$ defined by

$$\mu^T(B) = \mu(T^{-1}(B)) \quad \text{for all } B \in \mathcal{G}.$$

Intuitively, μ^T describes the distribution of the random element $T(\omega)$ when ω is distributed according to μ.

The relationship between integrals under the original and push-forward measures is given by the change of variables formula.

Theorem B.14 (Change of Variables Formula)**.** *Let $T : (\Omega, \mathcal{F}, \mu) \to (S, \mathcal{G})$ be measurable. For any measurable function $g : S \to \mathbb{R}$, g is integrable with respect to μ^T if and only if $g \circ T$ is integrable with respect to μ, and*

$$\int_S g(y)\, d\mu^T(y) = \int_\Omega g(T(\omega))\, d\mu(\omega).$$

Definition B.15 (Equivalence Relation)**.** An *equivalence relation* $\sim$ on a set X is a binary relation that satisfies three properties for all $a, b, c \in X$:

1. **Reflexivity:** $a \sim a$.
2. **Symmetry:** If $a \sim b$, then $b \sim a$.
3. **Transitivity:** If $a \sim b$ and $b \sim c$, then $a \sim c$.

Given an equivalence relation $\sim$ on a set X, the *equivalence class* of an element $x \in X$, denoted $[x]$, is the set of all elements in X equivalent to x:

$$[x] = \{y \in X : y \sim x\}.$$

The set of all equivalence classes is called the *quotient set* and denoted by $X/\sim$.

Equivalence relations allow us to define measurable structures on quotient spaces.

Definition B.16 (Quotient σ-algebra)**.** Let (X, Σ) be a measurable space and $\sim$ an equivalence relation on X. The *quotient σ-algebra* on the quotient space $X/\sim$, denoted by $\Sigma/\sim$, is defined as

$$\Sigma/\sim = \left\{B \subseteq X/\sim \,\middle|\, \pi^{-1}(B) \in \Sigma\right\},$$

where $\pi : X \to X/\sim$ is the canonical projection map $\pi(x) = [x]$.

This is the largest σ-algebra on $X/\sim$ making the projection π measurable.

B.2 Integration

B.2.1 The Standard Machinery

A common strategy in measure theory to prove a property $\mathfrak{p}$ for all measurable functions is the so-called "standard machine" or "approximation by simple functions". The steps are typically:

1. **Indicator Functions:** Prove that $\mathfrak{p}$ holds for indicator functions $\mathbb{1}_A$ for all measurable sets A.
2. **Simple Functions:** Extend the result to non-negative simple functions $s = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}$ by linearity.
3. **Non-negative Measurable Functions:** Use the fact that any non-negative measurable function f is the limit of an increasing sequence of non-negative simple functions $s_n \uparrow f$. Prove that $\mathfrak{p}$ is preserved under this limit (often using the Monotone Convergence Theorem).

4. **General Measurable Functions:** For a general measurable function f, write $f = f^+ - f^-$ where $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$. Extend the result by linearity, provided integrability conditions are met.

Key theorems supporting this machinery include:

Theorem B.17 (Monotone Class Theorem). *Let $\mathcal{A}$ be an algebra of sets generating a σ-algebra $\mathcal{F}$. Let $\mathcal{M}$ be a collection of subsets of Ω that is a monotone class (i.e., closed under countable increasing unions and countable decreasing intersections). If $\mathcal{A} \subseteq \mathcal{M}$, then $\mathcal{F} \subseteq \mathcal{M}$.*

Theorem B.18 (Monotone Convergence Theorem). *If $\{f_n\}$ is a sequence of non-negative measurable functions such that $f_n \uparrow f$ pointwise, then*

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Lemma B.19 (Fatou's Lemma). *If $\{f_n\}$ is a sequence of non-negative measurable functions, then*

$$\int \liminf_{n\to\infty} f_n \, d\mu \leqslant \liminf_{n\to\infty} \int f_n \, d\mu.$$

Theorem B.20 (Dominated Convergence Theorem). *Let $\{f_n\}$ be a sequence of measurable functions converging pointwise to f. If there exists an integrable function g such that $|f_n| \leqslant g$ for all n, then f is integrable and*

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

B.2.2 Function spaces

Definition B.21 ($\mathcal{L}^p$ spaces). Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. For $1 \leqslant p < \infty$, let $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ denote the set of all measurable functions $f : \Omega \to \mathbb{R}$ such that

$$\|f\|_p := \left(\int_\Omega |f|^p \, d\mu \right)^{1/p} < \infty.$$

Similarly, $\mathcal{L}^\infty(\Omega, \mathcal{F}, \mu)$ consists of all essentially bounded measurable functions, i.e., those for which there exists a constant C such that $|f(\omega)| \leqslant C$ for almost all ω. The essential supremum is defined as:

$$\|f\|_\infty := \inf\{C \geqslant 0 : |f(\omega)| \leqslant C \text{ for } \mu\text{-almost all } \omega\}.$$

The quantity $\|\cdot\|_p$ satisfies most properties of a norm (non-negativity, homogeneity, triangle inequality), but it is only a *semi-norm* on $\mathcal{L}^p$, because $\|f\|_p = 0$ implies $f = 0$

only almost everywhere (not everywhere). To obtain a Banach space, we must identify functions that are equal almost everywhere.

Definition B.22 (L^p spaces)**.** We define an equivalence relation $\sim$ on $\mathcal{L}^p$ by $f \sim g$ if and only if $f = g$ μ-almost everywhere. The L^p *space* is the quotient space of equivalence classes:

$$L^p(\Omega, \mathcal{F}, \mu) := \mathcal{L}^p(\Omega, \mathcal{F}, \mu)/\sim .$$

Elements of L^p are equivalence classes $[f]$, but it is standard practice to abuse notation and refer to them as functions f.

Equipped with the norm $\|[f]\|_p := \|f\|_p$, the space L^p becomes a Banach space (a complete normed vector space).

Important special cases include:

- $L^p(\mathbb{R}^d)$: When $\Omega = \mathbb{R}^d$ equipped with the Lebesgue measure.
- $L^p([0,1])$: The space of functions on the unit interval square-integrable with respect to Lebesgue measure. This is a standard setting for functional analysis.
- ℓ^p: When μ is the counting measure on $\mathbb{N}$, the space is the set of sequences (x_n) with $\sum |x_n|^p < \infty$.
- $L^2(\mu)$: For $p = 2$, the space is a Hilbert space with inner product $\langle f, g \rangle = \int fg\, d\mu$.

B.2.3 Change of measure

Let μ be a measure on $(\Omega, \mathcal{F})$ and let $f : \Omega \to [0, \infty]$ be a non-negative measurable function. We can define a new measure ν on $(\Omega, \mathcal{F})$ by setting

$$\nu(A) = \int_A f\, d\mu \quad \text{for all } A \in \mathcal{F}.$$

It is a standard exercise in measure theory to verify that ν indeed satisfies the properties of a measure.

Definition B.23 (Probability Density)**.** If the function f is non-negative and the induced measure ν satisfies $\nu(\Omega) = 1$ (i.e., ν is a probability measure), then f is called a *probability density* of ν with respect to the reference measure μ.

The relationship between ν and μ constructed above implies a specific property called absolute continuity.

Definition B.24 (Absolute Continuity)**.** Let ν and μ be two measures on a measurable space $(\Omega, \mathcal{F})$. We say ν is *absolutely continuous* with respect to μ (denoted $\nu \ll \mu$) if for all $A \in \mathcal{F}$,

$$\mu(A) = 0 \implies \nu(A) = 0.$$

The fundamental result connecting these concepts is the Radon-Nikodym theorem, which states that under mild conditions, absolute continuity is sufficient to guarantee the existence of a density.

Theorem B.25 (Radon-Nikodym Theorem)**.** *Let* ν *and* μ *be two measures on a measurable space* $(\Omega, \mathcal{F})$*, and assume that* μ *is* σ*-finite. If* $\nu \ll \mu$*, then there exists a non-negative measurable function* $f : \Omega \to [0, \infty)$ *such that for all* $A \in \mathcal{F}$*,*

$$\nu(A) = \int_A f \, d\mu.$$

The function f *is unique up to a set of* μ*-measure zero. We call* f *the* Radon-Nikodym derivative *or* density *of* ν *with respect to* μ*, and denote it by* $f = \frac{d\nu}{d\mu}$.

The next theorem provides a characterization of sufficient statistics (Definition 1.9). The theorem provides the measure-theoretic foundation for the Factorization Theorem (Theorem 1.12) encountered in the main text. Its proof is quite involved and is omitted here, but one can find it in Halmos and Savage 1949.

Theorem B.26 (Halmos–Savage)**.** *Let* $\mathcal{P}$ *be a family of probability measures dominated by a* σ*-finite measure. A statistic* T *is sufficient for* $\mathcal{P}$ *if and only if for all* $P, Q \in \mathcal{P}$*, the likelihood ratio* dP/dQ *admits a* $\sigma(T)$*-measurable version.*

B.3 Joint distributions

B.3.1 Product measures and independence

Given two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, the *product* σ*-algebra*, denoted $\mathcal{F}_1 \otimes \mathcal{F}_2$, is the σ-algebra on $\Omega_1 \times \Omega_2$ generated by the collection of measurable rectangles $\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$.

If μ_1 and μ_2 are σ-finite measures on $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ respectively, there exists a unique measure $\mu = \mu_1 \otimes \mu_2$ on the product space such that

$$\mu(A \times B) = \mu_1(A)\mu_2(B) \quad \text{for all } A \in \mathcal{F}_1, B \in \mathcal{F}_2.$$

Definition B.27 (Independence)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space. Two events $A, B \in \mathcal{F}$ are *independent* if $P(A \cap B) = P(A)P(B)$. Two random variables $X : \Omega \to \mathcal{X}$ and $Y : \Omega \to \mathcal{Y}$ are *independent* if for all $A \in \mathscr{X}$ and $B \in \mathscr{Y}$, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent.

In terms of joint distributions, independence means the joint distribution is the product measure of the marginals. That is, the joint law of (X, Y) is $P_{(X,Y)} = P_X \otimes P_Y$.

Definition B.28 (i.i.d.)**.** A sequence of random variables $X_1, X_2, \ldots, X_n$ is *independent and identically distributed (i.i.d.)* if they are mutually independent and all have the same marginal distribution.

If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P$, their joint distribution on the product space $(\mathcal{X}^n, \mathscr{X}^{\otimes n})$ is the product measure $P^{\otimes n}$, defined inductively by $P^{\otimes 1} = P$ and $P^{\otimes (n+1)} = P^{\otimes n} \otimes P$.

B.3.2 Conditional probability and expectation

The definition of conditional probability is based on the concept of conditional expectation.

Definition B.29 (Conditional Expectation)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra, and X an integrable random variable (i.e., $E|X| < \infty$). The *conditional expectation* of X given $\mathcal{G}$, denoted $E[X \mid \mathcal{G}]$, is the equivalence class of $\mathcal{G}$-measurable random variables Z such that

$$\int_G Z\, dP = \int_G X\, dP \quad \text{for all } G \in \mathcal{G}.$$

The existence and uniqueness (up to almost sure equivalence) of Z are guaranteed by the Radon-Nikodym theorem.

Theorem B.30 (Existence and Uniqueness of Conditional Expectation)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra, and X an integrable random variable. Then there exists a unique (up to almost sure equivalence) $\mathcal{G}$-measurable random variable Z such that*

$$\int_G Z\, dP = \int_G X\, dP \quad \text{for all } G \in \mathcal{G}.$$

With this tool, we can rigorously define the probability of an event given partial information.

Definition B.31 (Conditional Probability)**.** The *conditional probability* of an event $A \in \mathcal{F}$ given a sub-σ-algebra $\mathcal{G}$, denoted $P(A \mid \mathcal{G})$, is defined as the conditional expectation of the indicator function of A:

$$P(A \mid \mathcal{G}) := E[\mathbb{1}_A \mid \mathcal{G}].$$

When conditioning on a random variable Y, we mean conditioning on the σ-algebra generated by Y, i.e., $E[X \mid Y] := E[X \mid \sigma(Y)]$.

Conditional expectations satisfy a generalized version of Bayes' theorem.

Theorem B.32 (Abstract Bayes Formula). *Let P and Q be probability measures on $(\Omega, \mathcal{F})$ such that $P \ll Q$, and let $L = dP/dQ$ be the Radon-Nikodym derivative. For any sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and any P-integrable random variable f,*

$$E_P[f \mid \mathcal{G}] = \frac{E_Q[fL \mid \mathcal{G}]}{E_Q[L \mid \mathcal{G}]} \quad P\text{-a.s.}$$

Often, we want to view the conditional probability $P(\cdot \mid \mathcal{G})(\omega)$ as a probability measure on $(\Omega, \mathcal{F})$ for each fixed ω. This is not guaranteed by the general definition (due to null sets for each A). However, it is possible in sufficiently "nice" spaces.

A crucial property relating measurability with respect to a random variable and functions of that random variable is given by the Doob-Dynkin Lemma.

Lemma B.33 (Doob-Dynkin Lemma). *Let $X : \Omega \to S$ be a measurable map into a measurable space $(S, \mathcal{S})$. A function $Y : \Omega \to \mathbb{R}$ is $\sigma(X)$-measurable if and only if there exists a measurable function $g : S \to \mathbb{R}$ such that $Y = g(X)$.*

This lemma implies that $E[Z \mid X] = g(X)$ for some measurable function g. Specifically, if Y is $\sigma(X)$-measurable, it is a function of X.

Under certain conditions, conditional probabilities can be realized as a kernel that is a measure for each fixed ω.

Definition B.34 (Regular Conditional Probability). Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra. A *regular conditional probability* is a function $\kappa : \Omega \times \mathcal{F} \to [0, 1]$ such that:

1. For each $\omega \in \Omega$, $\kappa(\omega, \cdot)$ is a probability measure on $(\Omega, \mathcal{F})$.
2. For each $A \in \mathcal{F}$, $\omega \mapsto \kappa(\omega, A)$ is a version of $P(A \mid \mathcal{G})$.

Regular conditional probabilities are guaranteed to exist when Ω is a standard Borel space (e.g. a Polish space (see Definition A.15 in Appendix A) equipped with its Borel σ-algebra).

Theorem B.35 (Existence of Regular Conditional Probabilities). *Let $(\Omega, \mathcal{F}, P)$ be a probability space where Ω is a Polish space and $\mathcal{F} = \mathcal{B}(\Omega)$ is its Borel σ-algebra. For any sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, there exists a regular conditional probability given $\mathcal{G}$.*

A related concept is the Markov kernel, which generalizes the idea of a transition matrix.

Definition B.36 (Markov Kernel). Let $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$ be measurable spaces. A *Markov kernel* (or probability kernel) from $(X, \mathscr{X})$ to $(Y, \mathscr{Y})$ is a function $K : X \times \mathscr{Y} \to [0, 1]$ such that:

1. For each $x \in X$, the map $B \mapsto K(x, B)$ is a probability measure on $(Y, \mathscr{Y})$.

2. For each $B \in \mathscr{Y}$, the map $x \mapsto K(x, B)$ is $\mathscr{X}$-measurable.

Markov kernels are used to model random mappings where the output distribution depends on the input, such as in conditional distributions $P(Y \in B \mid X = x)$.

Finally, we state the version of Bayes' rule for densities, which is the most common form used in statistical inference.

Theorem B.37 (Bayes' Rule for Densities). *Let Θ and $\mathcal{X}$ be random variables taking values in measurable spaces $(\Omega_\Theta, \mathscr{F}_\Theta)$ and $(\Omega_\mathcal{X}, \mathscr{F}_\mathcal{X})$, respectively. Suppose the joint distribution of $(\Theta, \mathcal{X})$ is dominated by a product measure $\nu \otimes \mu$, with joint density $p(\theta, x)$. Then the conditional distribution of Θ given $\mathcal{X} = x$ has density (with respect to ν):*

$$p(\theta \mid x) = \frac{p(\theta, x)}{\int_{\Omega_\Theta} p(\vartheta, x)\, d\nu(\vartheta)},$$

provided the denominator is positive and finite. In the common case where $p(\theta, x) = p(x \mid \theta)\pi(\theta)$ (likelihood $\times$ prior), this becomes the familiar form:

$$p(\theta \mid x) = \frac{p(x \mid \theta)\pi(\theta)}{\int p(x \mid \vartheta)\pi(\vartheta)\, d\nu(\vartheta)}.$$

B.4 Concentration of measure

Lemma B.38 (Jensen's Inequality). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, X an integrable real-valued random variable, and $\varphi : \mathbb{R} \to \mathbb{R}$ a convex function. Then*

$$\varphi(E[X]) \leqslant E[\varphi(X)].$$

If φ is strictly convex, then equality holds if and only if X is constant almost surely.

Lemma B.39 (Markov's inequality). *If $X \geqslant 0$, then for any $a > 0$,*

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}.$$

Proof. Note that $a \cdot \mathbb{1}X \geqslant a \leqslant X$. Taking expectations gives $a \cdot \mathbb{P}(X \geqslant a) \leqslant \mathbb{E}[X]$. □

The following concentration inequalities are immediate consequences.

Lemma B.40 (Chebyshev's inequality). *If $\mathrm{Var}(X) < \infty$, then for any $k > 0$,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant k) \leqslant \frac{\mathrm{Var}(X)}{k^2}.$$

Proof. Apply Markov's inequality to $(X - \mathbb{E}[X])^2$ with threshold k^2. □

Lemma B.41 (Chernoff's bound)**.** *For any random variable X and any $a \in \mathbb{R}$,*

$$\mathbb{P}(X \geqslant a) \leqslant \inf_{t>0} e^{-ta}\mathbb{E}[e^{tX}].$$

Proof. For any $t > 0$, the event $\{X \geqslant a\}$ implies $\{e^{tX} \geqslant e^{ta}\}$. Apply Markov's inequality to e^{tX} and take the infimum over $t > 0$. □

B.5 Transforms

Definition B.42 (Laplace Transform)**.** Let μ be a finite measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. The *Laplace transform* of μ is the function $\psi : \mathbb{R}^k \to \mathbb{R}$ defined by

$$\psi(t) = \int_{\mathbb{R}^k} e^{\langle t,x\rangle}\, d\mu(x),$$

provided the integral exists.

The Laplace transform is a powerful tool for characterizing measures. A key property is its uniqueness:

Theorem B.43 (Uniqueness of Laplace Transform)**.** *Let μ and ν be two finite measures on $\mathbb{R}^k$. If their Laplace transforms agree on an open set containing the origin, then $\mu = \nu$.*

Proof Sketch. We sketch the argument for $k = 1$ and compact support. Suppose μ and ν are supported on a compact interval $[a, b]$. The Laplace transform condition implies

$$\int_a^b e^{tx}\, d\mu(x) = \int_a^b e^{tx}\, d\nu(x)$$

for all t in a neighborhood of 0. By analyticity, this equality extends to all $t \in \mathbb{R}$. By linearity,

$$\int_a^b P(e^x)\, d\mu(x) = \int_a^b P(e^x)\, d\nu(x)$$

for any polynomial P. The algebra of functions of the form $x \mapsto P(e^x)$ separates points on $[a, b]$ and vanishes at no point. By the Stone-Weierstrass theorem, such functions are dense in the space of continuous functions $C([a, b])$ with respect to the uniform norm.

Thus, for any continuous function f, $\int f\, d\mu = \int f\, d\nu$. Since measures on Borel σ-algebras are determined by their integrals against continuous functions (Riesz Representation Theorem), we conclude $\mu = \nu$. The extension to non-compact support requires more careful analysis involving truncation or compactification, but the core idea remains the density of exponential families in function spaces. □

This uniqueness property extends to signed measures. If μ is a signed measure with $\int e^{\langle t,x\rangle}\, d\mu(x) = 0$ for all t in an open set, then μ is the zero measure. This fact is crucial for proving completeness of exponential families.

Another important transform is the characteristic function, which similarly provides a powerful tool for characterizing measures.

Definition B.44 (Characteristic Function)**.** Let μ be a finite measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. The *characteristic function* of μ is the function $\phi : \mathbb{R}^k \to \mathbb{C}$ defined by

$$\phi(t) = \int_{\mathbb{R}^k} e^{i\langle t,x\rangle}\, d\mu(x),$$

where $i = \sqrt{-1}$.

Unlike the Laplace transform, the characteristic function is always defined for any finite measure (since $|e^{i\langle t,x\rangle}| = 1$ is bounded). It also uniquely determines the measure.

Theorem B.45 (Uniqueness of Characteristic Functions)**.** *Let μ and ν be two finite measures on $\mathbb{R}^k$. If their characteristic functions agree, i.e., $\phi_\mu(t) = \phi_\nu(t)$ for all $t \in \mathbb{R}^k$, then $\mu = \nu$.*

This theorem is a direct consequence of the Fourier Inversion Theorem. Since the characteristic function is essentially the Fourier transform of the measure, and the Fourier transform is injective, the measure is uniquely determined.