Probability spaces

Definition

A *probability space* is a measure space \((\Omega, \mathcal{F}, \text{Prob})\) with \(\text{Prob}\) a positive measure of mass 1.

- \(\Omega\) is called the *sample space*, and \(\omega \in \Omega\) are called outcomes.
- \(\mathcal{F}\), a \(\sigma\)-algebra, is called the *event space*, and \(A \in \mathcal{F}\) are called *events*. 

**Definition**

A collection of sets \( \mathcal{I} \) is a *semialgebra* if

- If \( S, T \in \mathcal{I} \) then \( S \cap T \in \mathcal{I} \)
- If \( S \in \mathcal{I} \) then \( S^c \) is the finite disjoint union of sets of \( \mathcal{I} \).

**Example**

The empty set together with those sets

\[ (a_1, b_1] \times \cdots \times (a_d, b_d] \subset \mathbb{R}^d, \quad -\infty \leq a_i < b_i \leq \infty \]

form a semialgebra in \( \mathbb{R}^d \).
Algebras of sets

**Definition**

A collection of sets $\mathcal{I}$ is an *algebra* if it is closed under complements and intersections.

**Lemma**

*If $\mathcal{I}$ is a semialgebra, then $\mathcal{I}'$, given by finite disjoint unions from $\mathcal{I}$, is an algebra.*

**Definition**

A $\sigma$-algebra of sets is an algebra which is closed under countable unions.
Borel $\sigma$-algebra

**Definition**
Given a collection of subsets $A_\alpha \subset \Omega$, the *generated $\sigma$-algebra* $\sigma(\{A_\alpha\})$ is the smallest $\sigma$-algebra containing $\{A_\alpha\}$.

**Definition**
In the case that $\Omega$ has a topology $\mathcal{T}$ of open sets, the *Borel $\sigma$-algebra* is $\sigma(\mathcal{T})$. 
Borel $\sigma$-algebra

**Definition**

The product of measure spaces $(\Omega_i, \mathcal{F}_i), \ i = 1, \ldots, n$ is the set $\Omega = \Omega_1 \times \ldots \times \Omega_n$ with the $\sigma$-algebra $\mathcal{F}_1 \times \ldots \times \mathcal{F}_n = \sigma(\bigcup_{i=1}^n \mathcal{F}_i)$.

**Exercise**

Let $d \geq 1$. With the usual topologies, the Borel $\sigma$-algebra $\mathcal{B}_{\mathbb{R}^d}$ is equal to $\mathcal{B}_{\mathbb{R}} \times \ldots \times \mathcal{B}_{\mathbb{R}}$ ($d$ copies).
Dynkin’s $\pi - \lambda$ Theorem

Definition

A $\pi$-system is a collection $\mathcal{P}$ of sets closed under finite intersections. A $\lambda$-system is a collection $\mathcal{L}$ of sets satisfying the following:

1. $\Omega \in \mathcal{L}$
2. For any $A, B \in \mathcal{L}$ satisfying $A \subset B$, $B \setminus A \in \mathcal{L}$
3. If $A_1 \subset A_2 \subset \ldots$ is a sequence from $\mathcal{L}$ and $A = \bigcup_{i=1}^{\infty} A_i$ then $A \in \mathcal{L}$.
Lemma

Let $\mathcal{L}$ be a $\lambda$-system which is closed under intersection. Then $\mathcal{L}$ is a $\sigma$-algebra.

Proof.

- If $A \in \mathcal{L}$ then $A^c = \Omega \setminus A \in \mathcal{L}$.
- If $A, B \in \mathcal{L}$ then $A \cup B = (A^c \cap B^c)^c \in \mathcal{L}$.
- Thus, if $\{A_i\}_{i=1}^{\infty}$ is a sequence in $\mathcal{L}$, then for each $n$, $\bigcup_{i=1}^{n} A_i \in \mathcal{L}$, and hence $\bigcup_{i=1}^{\infty} A_i \in \mathcal{L}$. 
Dynkin’s $\pi - \lambda$ Theorem

Theorem (Dynkin’s $\pi - \lambda$ Theorem)

If $\mathcal{P} \subseteq \mathcal{L}$ with $\mathcal{P}$ a $\pi$-system and $\mathcal{L}$ a $\lambda$-system then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$. 
Dynkin’s $\pi - \lambda$ Theorem

Proof.

Let $\ell(\mathcal{P})$ be the smallest $\lambda$-system containing $\mathcal{P}$. We show that $\ell(\mathcal{P})$ is a $\sigma$-algebra.

- Let $A \in \ell(\mathcal{P})$ and define $L_A = \{ B : A \cap B \in \ell(\mathcal{P}) \}$.
- We check that $L_A$ is a $\lambda$-system.
  - $\Omega \in L_A$ since $A \in \ell(\mathcal{P})$
  - If $B, C \in L_A$ and $B \supset C$, then $A \cap (B - C) = (A \cap B) - (A \cap C) \in \ell(\mathcal{P})$.
  - If $B_1 \subset B_2 \subset \ldots$ is a sequence from $L_A$ with $B = \bigcup_{i=1}^{\infty} B_i$ then $B_1 \cap A \subset B_2 \cap A \subset \ldots$ has $B \cap A = \bigcup_{i=1}^{\infty} (B_i \cap A)$, and hence $B \cap A \in \ell(\mathcal{P})$ so $B \in L_A$.

- If $A \in \mathcal{P}$ then $L_A = \ell(\mathcal{P})$. Hence, if $B \in \ell(\mathcal{P})$ then $A \cap B \in \ell(\mathcal{P})$. But then this implies $L_B = \ell(\mathcal{P})$. It follows that for all $A, B \in \ell(\mathcal{P})$, $A \cap B \in \ell(\mathcal{P})$. 

**Measures**

**Definition**

A *positive measure on an algebra* $\mathcal{A}$ is a set function $\mu$ which satisfies

- $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{A}$
- If $A_i \in \mathcal{A}$ are disjoint and their union is in $\mathcal{A}$, then
  \[ \mu \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i). \]

If $\mu(\Omega) = 1$ then $\mu$ is a *probability measure*. 
A probability measure satisfies the following basic properties.

- (Monotonicity) If $A \subset B$ then $\text{Prob}(A) \leq \text{Prob}(B)$.
- (Sub-additivity) If $A \subset \bigcup_i A_i$ then $\text{Prob}(A) \leq \sum_i \text{Prob}(A_i)$
- (Continuity from below) If $A_1 \subset A_2 \subset \ldots$ and $A = \bigcup_i A_i$ then $\text{Prob}(A_i) \uparrow \text{Prob}(A)$
- (Continuity from above) If $A_1 \supset A_2 \supset \ldots$ and $A = \bigcap_i A_i$ then $\text{Prob}(A_i) \downarrow \text{Prob}(A)$. 
A probability space \((\Omega, \mathcal{F}, \text{Prob})\) is non-atomic if \(\text{Prob}(A) > 0\) implies that there exists \(B \in \mathcal{F}\) satisfying \(B \subset A\) and \(0 < \text{Prob}(B) < \text{Prob}(A)\).
An outer measure $\mu^*$ on a measurable space $(\Omega, \mathcal{F})$ is a set function $\mu^* : \mathcal{F} \rightarrow [0, \infty]$ satisfying

- $\mu^*(\emptyset) = 0$ and $\mu^*(A_1) \leq \mu^*(A_2)$ for any $A_1, A_2 \in \mathcal{F}$ with $A_1 \subset A_2$.
- $\mu^* \left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$ for any countable collection of sets $\{A_n\} \subset \mathcal{F}$. 
Definition

Given an outer measure $\mu^*$ on a measurable space $(\Omega, \mathcal{F})$, a set $A \in \mathcal{F}$ is measurable (in the sense of Carathéodory) if for each set $E \in \mathcal{F}$,

$$
\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).
$$
Theorem

Let $\mu^*$ be an outer measure on a measurable space $(\Omega, \mathcal{F})$. The subset $\mathcal{G}$ of $\mu^*$-measurable sets in $\mathcal{F}$ is a $\sigma$-algebra, and $\mu^*$ restricted to this subset is a measure.

See e.g. Royden pp.54–60.
An outer measure on \((\mathbb{R}, 2\mathbb{R})\) is given by

\[
\mu^* (A) = \inf \left\{ \sum_{i=1}^{\infty} b_i - a_i : A \subset \bigcup_{i=1}^{\infty} (a_i, b_i] \right\}.
\]

Lebesgue measure is obtained by restricting \(\mu^*\) to its measurable sets. The \(\sigma\)-algebra so obtained is larger than the Borel \(\sigma\)-algebra.
Theorem

Let $\mu$ be a $\sigma$-finite measure on an algebra $\mathcal{A}$. Then $\mu$ has a unique extension to $\sigma(\mathcal{A})$. 
Proof of uniqueness.

Let $\mu_1$ and $\mu_2$ be two extensions of $\mu$ to $\sigma(\mathcal{A})$. Let $A \in \mathcal{A}$ satisfy $\mu(A) < \infty$ and let

$$\mathcal{L} = \{B \in \sigma(\mathcal{A}) : \mu_1(A \cap B) = \mu_2(A \cap B)\}.$$  

We show that $\mathcal{L}$ is a $\lambda$-system. Since $\mathcal{A} \subset \mathcal{L}$ and $\mathcal{A}$ is a $\pi$-system, it then follows that $\mathcal{L} = \sigma(\mathcal{A})$. Uniqueness then follows on taking a sequence $\{A_n\}$ with $A_n \uparrow \Omega$ and $\mu(A_n) < \infty$. 
Proof of uniqueness.

To verify the $\lambda$-system property, observe

- $\Omega \in \mathcal{L}$
- If $B, C \in \mathcal{L}$ with $C \subset B$, then

$$\mu_1(A \cap (B - C)) = \mu_1(A \cap B) - \mu_1(A \cap C)$$

$$= \mu_2(A \cap B) - \mu_2(A \cap C) = \mu_2(A \cap (B - C)).$$

- If $B_n \in \mathcal{L}$ and $B_n \uparrow B$ then

$$\mu_1(A \cap B) = \lim_{n \to \infty} \mu_1(A \cap B_n) = \lim_{n \to \infty} \mu_2(A \cap B_n) = \mu_2(A \cap B).$$
Carathéodory’s Extension Theorem

Proof of existence.

Define set function $\mu^*$ on $\sigma(\mathcal{A})$ by

$$\mu^*(E) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) : E \subset \bigcup_{i=1}^{\infty} A_i, A_i \in \mathcal{A} \right\}.$$

Evidently $\mu^*(A) = \mu(A)$ for $A \in \mathcal{A}$. Also, $A \in \mathcal{A}$ is measurable, since for $F \in \sigma(\mathcal{A})$ and $\epsilon > 0$ there exists $\{B_i\}_{i=1}^{\infty}$ a sequence from $\mathcal{A}$ satisfying $\sum_i \mu(B_i) \leq \mu^*(F) + \epsilon$. Then

$$\mu(B_i) = \mu^*(B_i \cap A) + \mu^*(B_i \cap A^c)$$

$$\mu^*(F) + \epsilon \geq \sum_i \mu^*(B_i \cap A) + \sum_i \mu^*(B_i \cap A^c) \geq \mu^*(F \cap A) + \mu^*(F^c \cap A).$$

which gives the condition for measurability.
Carathéodory’s Extension Theorem

Proof of existence.

\( \mu^* \) satisfies the properties of an outer measure, since

- If \( E \subset F \) then \( \mu^*(E) \leq \mu^*(F) \)
- If \( F \subset \bigcup_i F_i \) is a countable union, then \( \mu^*(F) \leq \sum_i \mu^*(F_i) \).

The restriction of \( \mu^* \) to its measurable sets gives the required extension of \( \mu \).
Random variables

**Definition**

A real valued *random variable* on a measure space \((\Omega, \mathcal{F}, \text{Prob})\) is a function \(X: \Omega \rightarrow \mathbb{R}\) which is \(\mathcal{F}\)-measurable, that is, for each Borel set \(B \subset \mathbb{R}\),

\[
X^{-1}(B) = \{\omega: X(\omega) \in B\} \in \mathcal{F}.
\]

A *random vector* in \(\mathbb{R}^d\) is a measurable map \(X: \Omega \rightarrow \mathbb{R}^d\).

Given \(A \in \mathcal{F}\), the indicator function of \(A\) is a random variable,

\[
1_A(\omega) = \begin{cases} 
1 & \omega \in A \\
0 & \omega \notin A
\end{cases}.
\]
Random variables

**Theorem**

If $X_1, ..., X_n$ are random variables and $f : (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}) \to (\mathbb{R}, \mathcal{B})$ is measurable, then $f(X_1, ..., X_n)$ is a random variable.

**Theorem**

If $X_1, X_2, ...$ are random variables then $X_1 + X_2 + ... + X_n$ is a random variable, and so are

$$\inf_n X_n, \quad \sup_n X_n, \quad \lim \sup_n X_n, \quad \lim \inf_n X_n.$$ 

**Proof.**

Exercise, or see Durrett, pp. 14–15.
Distributions

Definition

The *distribution* of a random variable $X$ on $\mathbb{R}$ is the probability measure $\mu$ on $(\mathbb{R}, \mathcal{B})$ defined by

$$\mu(A) = \text{Prob}(X \in A).$$

The *distribution function* of $X$ is

$$F(x) = \text{Prob}(X \leq x).$$
Distributions

Theorem

Any distribution function $F$ has the following properties:

1. $F$ is nondecreasing.
2. $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$.
3. $F$ is right continuous, that is, $\lim_{y \downarrow x} F(y) = F(x)$.
4. If $F(x^-) = \lim_{y \uparrow x} F(y)$ then $F(x^-) = \text{Prob}(X < x)$.
5. $\text{Prob}(X = x) = F(x) - F(x^-)$.

Furthermore, any function satisfying the first three items is the distribution function of a random variable.
Proof.

All of the forward claims are straightforward. For the reverse claim, let $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}$ and set Prob to be Lebesgue measure. Define 

$$X(\omega) = \sup\{y : F(y) < \omega\}.$$ 

Then 

$$\{\omega : X(\omega) \leq x\} = \{\omega : \omega \leq F(x)\},$$

which follows by the right-continuity of $F$. Hence $\text{Prob}(X \leq x) = F(x).$
Distributions

Definition

If $X$ and $Y$ induce the same distribution $\mu$ on $\mathbb{R}$, we say $X$ and $Y$ are equal in distribution. We write $X =_{d} Y$.

Definition

When the distribution function $F(x) = \text{Prob}(X \leq x)$ has the form

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

we say that $X$ has density function $f$. 
Example distributions

- **Uniform distribution on (0,1).** Density $f(x) = 1$ for $x \in (0, 1)$ and 0 otherwise.

- **Exponential distribution with rate $\lambda$.** Density $f(x) = \lambda e^{-\lambda x}$ for $x > 0$, 0 otherwise.

- **Standard normal distribution.** Density $f(x) = \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}$. 

Example distributions

- Uniform distribution on the Cantor set. Define distribution function $F$ by $F(x) = 0$ for $x \leq 0$, $F(x) = 1$ for $x \geq 1$, $F(x) = \frac{1}{2}$ for $x \in \left[\frac{1}{3}, \frac{2}{3}\right]$, $F(x) = \frac{1}{4}$ for $x \in \left[\frac{1}{9}, \frac{2}{9}\right]$, $F(x) = \frac{3}{4}$ for $x \in \left[\frac{7}{9}, \frac{8}{9}\right]$,....

- Point mass at 0. The distribution function has $F(x) = 0$ for $x < 0$, $F(x) = 1$ for $x \geq 0$.

- Lognormal distribution. Let $X$ be a standard Gaussian variable. $\exp(X)$ is lognormal.

- Chi-square distribution. Let $X$ be a standard Gaussian variable. $X^2$ has a chi-squared distribution.
Example distributions on $\mathbb{Z}$

- Bernoulli distribution, parameter $p$. $\text{Prob}(X = 1) = p$, $\text{Prob}(X = 0) = 1 - p$.

- Poisson distribution, parameter $\lambda$. $X$ is supported on $\mathbb{Z}$ and $\text{Prob}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

- Geometric distribution, success probability $p \in (0, 1)$. $X$ is supported on $\mathbb{Z}$ and $\text{Prob}(X = k) = p(1 - p)^{k-1}$, for $k = 1, 2, ...$. 
The Lebesgue integral against a $\sigma$-finite measure is defined as usual for:

1. Simple functions
2. Bounded functions
3. Nonnegative functions
4. General functions
Theorem (Jensen’s inequality)

Let $\phi$ be convex on $\mathbb{R}$. If $\mu$ is a probability measure, and $f$ and $\phi(f)$ are integrable then

$$\phi \left( \int f \, d\mu \right) \leq \int \phi(f) \, d\mu.$$
Jensen’s inequality

Proof.

Let \( c = \int f \, d\mu \) and let \( \ell(x) = ax + b \) be a linear function which satisfies \( \ell(c) = \phi(c) \) and \( \phi(x) \geq \ell(x) \). Thus

\[
\int \phi(f) \, d\mu \geq \int (af + b) \, d\mu = \phi \left( \int f \, d\mu \right).
\]
Hölder’s inequality

**Theorem**

If $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q.$$
Hölder's inequality

Proof.

We may assume that \( \|f\|_p > 0 \) and \( \|g\|_q > 0 \), since otherwise both sides vanish. Dividing both sides by \( \|f\|_p \|g\|_q \), we may assume that \( \|f\|_p = \|g\|_q = 1 \).

For fixed \( y \geq 0 \),

\[
\phi(x) = \frac{x^p}{p} + \frac{y^q}{q} - xy
\]

has a minimum in \( x \geq 0 \) at \( x_0 = y^{\frac{1}{p-1}} \) and \( x_0^p = y^{\frac{p}{p-1}} = y^q \), so \( \phi(x_0) = 0 \).

Thus \( xy \leq \frac{x^p}{p} + \frac{y^q}{q} \) in \( x, y \geq 0 \). The claim follows by setting \( x = |f|, y = |g| \) and integrating. \( \square \)
**Bounded convergence theorem**

**Definition**

We say that $f_n \to f$ in measure if, for any $\epsilon > 0$,

$$
\mu(\{x : |f_n(x) - f(x)| > \epsilon\}) \to 0
$$

as $n \to \infty$.

**Theorem (Bounded convergence theorem)**

Let $E$ be a set with $\mu(E) < \infty$. Suppose $f_n$ vanishes on $E^c$, $|f_n(x)| \leq M$, and $f_n \to f$ in measure. Then

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

Let $\epsilon > 0$, $G_n = \{x : |f_n(x) - f(x)| < \epsilon\}$ and $B_n = E - G_n$. Thus

$$\left| \int f d\mu - \int f_n d\mu \right| \leq \int |f - f_n| d\mu$$

$$= \int_{G_n} |f - f_n| d\mu + \int_{B_n} |f - f_n| d\mu \leq \epsilon \mu(E) + 2M \mu(B_n).$$

Since $f_n \rightarrow f$ in measure, $\mu(B_n) \rightarrow 0$. The proof follows on letting $\epsilon \downarrow 0$. 

\[\blacksquare\]
Fatou’s lemma

Lemma (Fatou’s lemma)

If \( f_n \geq 0 \) then

\[
\liminf_{n \to \infty} \int f_n \, d\mu \geq \int \left( \liminf_{n \to \infty} f_n \right) \, d\mu.
\]
Fatou’s lemma

Proof.

Let $g_n(x) = \inf_{m \geq n} f_m(x)$, and note that

$$g_n(x) \uparrow g(x) = \liminf_{n \to \infty} f_n(x).$$

It suffices to verify that $\lim_{n \to \infty} \int g_n d\mu \geq \int g d\mu$. To do so, let $E_m \uparrow \Omega$ be sets of finite measure. For each fixed $m$, as $n \to \infty$,

$$\int g_n d\mu \geq \int_{E_m} g_n \wedge m d\mu \to \int_{E_m} g \wedge m d\mu.$$

Letting $m \to \infty$ proves the result.
Theorem (Monotone convergence theorem)

If $f_n \geq 0$ and $f_n \uparrow f$ then

$$\int f_n d\mu \uparrow \int f d\mu.$$
Monotone convergence theorem

Proof.

By Fatou’s lemma, \( \lim_{n \to \infty} \int f_n \, d\mu \geq \int f \, d\mu \). The reverse inequality is immediate.
**Theorem (Dominated convergence theorem)**

If \( f_n \to f \) a.e., \(|f_n| \leq g \) for all \( n \) and \( g \) is integrable, then \( \int f_n \, d\mu \to \int f \, d\mu \).
**Dominated convergence theorem**

**Proof.**

Since $f_n + g \geq 0$, Fatou’s lemma gives

$$\liminf_{n \to \infty} \int (f_n + g) d\mu \geq \int (f + g) d\mu.$$ 

Thus $\liminf_{n \to \infty} \int f_n d\mu \geq \int f d\mu$. To prove the limit, replace $f_n$ with $-f_n$. 

Expected value

**Definition**

Let $X$ be a random variable on $(\Omega, \mathcal{F}, \text{Prob})$, and write $X = X^+ + X^−$ in a positive and negative part. The *expected value* of $X^+$ is $E[X^+] = \int X^+ dP$, similarly $X^−$. If either $E[X^+]$ or $E[X^−]$ is finite we say $E[X]$ exists and its value is

$$E[X] = E[X^+] + E[X^−].$$

$E[X]$ is also called the mean, $\mu$. 
### Expected value

**Theorem**

Suppose $X_n \to X$ a.s. Let $g$ and $h$ be continuous functions on $\mathbb{R}$ satisfying

- $g \geq 0$ and $g(x) \to \infty$ as $|x| \to \infty$
- $|h(x)|/g(x) \to 0$ as $|x| \to \infty$
- There exists $K \geq 0$ such that $E[g(X_n)] \leq K$ for all $n$.

Then $E[h(X_n)] \to E[h(X)]$ as $n \to \infty$.

A common application of this theorem takes $h(x) = x$ and $g(x) = |x|^p$ for some $p > 1$. 
Expected value

Proof.
The proof method is an example of truncation.

- Assume w.l.o.g. that $h(0) = 0$.
- Let $M > 0$ be such that $\text{Prob}(X = M) = 0$ and $g(x) > 0$ for $|x| > M$.
- Define $\overline{Y} = Y \mathbf{1}_{(|Y| \leq M)}$. By bounded convergence, $E[h(X_n)] \to E[h(X)]$. 


Bob Hough
Math 639: Lecture 1
January 24, 2017
47 / 54
Proof.

- Use

\[
|E[h(Y)] - E[h(Y)]| \leq E[|h(Y) - h(Y)|] = E[|h(Y)|1_{|Y|>M}] \leq \epsilon_M E[g(Y)]
\]

where \( \epsilon_M = \sup \{|h(x)| \cdot g(x) : |x| > M\} \).

- Thus \( |E[h(X_n)] - E[h(X_n)]| \leq K\epsilon_M \). Also,

\[
E[g(X)] \leq \liminf_{n \to \infty} E[g(X_n)] \leq K
\]

so \( |E[h(X)] - E[h(X)]| \leq K\epsilon_M \).
Expected value

Proof.

It follows from the triangle inequality that

\[ |E[h(X_n)] - E[h(X)]| \leq 2K\epsilon_M + |E[h(\overline{X}_n)] - E[h(\overline{X})]|. \]

Letting first \( n \), then \( m \) tend to infinity proves the claim.
Theorem

Let $X$ be a random element of $(S, \mathcal{S})$ with distribution $\mu$, that is, $\mu(A) = \text{Prob}(X \in A)$. If $f$ is measurable from $(S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B})$ and is such that $f \geq 0$ or $E[|f(X)|] < \infty$, then

$$E[f(X)] = \int_S f(y) \mu(dy).$$
Proof.

- If $B \in \mathcal{S}$ and $f = 1_B$ then

$$E[1_B(X)] = \text{Prob}(X \in B) = \mu(B) = \int_S 1_B(y)\mu(dy).$$

- The equality thus holds for simple functions by linearity.
- The equality holds for non-negative functions $f$ by taking a sequence of simple functions $f_n \uparrow f$ and applying monotone convergence.
- The equality holds for general $f$ by linearity again.
Definition

Let $X$ be a random variable which is square integrable. The variance of $X$ is

$$\text{Var}(X) = E[X^2] - E[X]^2$$

and the standard deviation is $\sigma = \text{Var}(X)^{\frac{1}{2}}$. 
Markov’s inequality

**Theorem**

Let $X \geq 0$ be a non-negative random variable with finite mean $\mu$. Then for all $\lambda \geq 1$,

$$\text{Prob}(X > \lambda \mu) \leq \frac{1}{\lambda}.$$ 

**Proof.**

The result holds if $\mu = 0$, so assume otherwise. Write

$$\lambda \mu \text{Prob}(X > \lambda \mu) \leq E[X 1_{X > \lambda \mu}] \leq E[X] = \mu$$

to conclude.
Chebyshev’s inequality

**Theorem**

Let $X$ be a square-integrable random variable with mean $\mu$ and standard deviation $\sigma$. Then for all $\lambda \geq 1$,

$$\text{Prob}(|X - \mu| > \lambda \sigma) \leq \frac{1}{\lambda^2}.$$ 

**Proof.**

Apply Markov’s inequality to $(X - \mu)^2$. 

□