Probability Review

Thomas Lonon

Division of Financial Engineering
Stevens Institute of Technology

May 17, 2016

©2016 The Trustees of the Stevens Institute of Technology

Outline

©2016 The Trustees of the Stevens Institute of Technology

Sigma Algebras

Definition 1.1.1: Let Ω be a nonempty set, and let $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a σ-algebra provided that:

1. The empty set belongs to $\mathcal{F}$
2. Whenever a set A belongs to $\mathcal{F}$, its complement A^c also belongs to $\mathcal{F}$
3. Whenever a sequence of sets $A_1, A_2, \ldots$ belongs to $\mathcal{F}$, their union $(\cup_{i=1}^{\infty} A_i)$ also belongs to $\mathcal{F}$

[1]

©2016 The Trustees of the Stevens Institute of Technology

Probability Measure

Definition 1.1.2: Let Ω be a nonempty set, and let $\mathcal{F}$ be a σ-algebra of subsets of Ω. A probability measure $\mathbb{P}$ is a function that, to every set $A \in \mathcal{F}$, assigns a number in $[0, 1]$, called the probability of A and written $\mathbb{P}(A)$. We require:

1. $\mathbb{P}(\Omega) = 1$
2. Whenever $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{F}$, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.[1]

Definition 1.1.5: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If a set $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, we say that the event A occurs almost surely.[1]

©2016 The Trustees of the Stevens Institute of Technology

Random Variable

Definition 1.2.1: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable is a real-valued function X defined on Ω with the property that for every Borel subset B of $\mathbb{R}$, the subset of Ω given by

$$\{X \in B\} = \{\omega \in \Omega; X(\omega) \in B\}$$

is in the σ-algebra $\mathcal{F}$.[1]

©2016 The Trustees of the Stevens Institute of Technology

Theorem 1.3.1: Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

- If X takes only finite many values $y_0, y_1, \ldots, y_n$, then

$$\int_\Omega X(\omega) d\mathbb{P}(\omega) = \sum_{k=0}^{n} y_k \mathbb{P}\{X = y_k\}$$

- **Integrability:** The random variable X is integrable if and only if

$$\int_\Omega |X(\omega)| d\mathbb{P}(\omega) < \infty$$

©2016 The Trustees of the Stevens Institute of Technology

Now let Y be another random variable on $(\Omega, \mathcal{F}, \mathbb{P})$

- **Comparison:** If $X \leq Y$ almost surely and if $\int_\Omega X(\omega)d\mathbb{P}(\omega)$ and $\int_\Omega Y(\omega)d\mathbb{P}(\omega)$ are defined, then

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) \leq \int_\Omega Y(\omega)d\mathbb{P}(\omega)$$

  In particular, if $X = Y$ almost surely and one of the integrals is defined, then they are both defined and

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) = \int_\Omega Y(\omega)d\mathbb{P}(\omega)$$

- **Linearity:** If α and β are real constants and X and Y are integrable, or if α and β are nonnegative constants and X and Y are nonnegative, then

$$\int_\Omega(\alpha X(\omega) + \beta Y(\omega))d\mathbb{P}(\omega) = \alpha \int_\Omega X(\omega)d\mathbb{P}(\omega) + \beta \int_\Omega Y(\omega)d\mathbb{P}(\omega)$$

[1]

©2016 The Trustees of the Stevens Institute of Technology

Theorem 1.3.4: Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

- If X takes only finitely many values $x_0, x_1, \ldots, x_n$, then

$$\mathbb{E}[X] = \sum_{k=0}^{n} x_k \mathbb{P}\{X = x_k\}$$

  In particular, if Ω is finite, then

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

- **Integrability:** The random variable X is integrable if and only if

$$\mathbb{E}[|X|] < \infty$$

©2016 The Trustees of the Stevens Institute of Technology

Now let Y be another random variable on $(\Omega, \mathcal{F}, \mathbb{P})$

- **Comparison:** If $X \leq Y$ almost surely and X and Y are integrable or almost surely nonnegative, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$. In particular, if $X = Y$ almost surely and one of the random variables is integrable or almost surely nonnegative, then they are both integrable or almost surely nonnegative, respectively, and $\mathbb{E}[X] = \mathbb{E}[Y]$
- **Linearity:** If α and β are real constants and X and Y are integrable or if α and β are nonnegative constants and X and Y are nonnegative, then

$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$$

- **Jensen's Inequality:** If φ is a convex, real-valued function defined on $\mathbb{R}$, and if $\mathbb{E}[X] < \infty$, then

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

[1]

©2016 The Trustees of the Stevens Institute of Technology

Expectation

For Discrete:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

$$\mathbb{E}[g(X)] = \sum_{\omega \in \Omega} g(X(\omega))\mathbb{P}(\omega)$$

For Continuous:

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$$

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega)) d\mathbb{P}(\omega)$$

©2016 The Trustees of the Stevens Institute of Technology

Change of Measure

Let us define the random variable

$$Z(\omega) = \frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}$$

but this is better written (to avoid dividing by 0) as

$$Z(\omega)\mathbb{P}(\omega) = \tilde{\mathbb{P}}(\omega)$$

©2016 The Trustees of the Stevens Institute of Technology

Theorem 1.6.1: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be an almost surely nonnegative random variable with $\mathbb{E}[Z] = 1$. For $A \in \mathcal{F}$, define

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega)$$

Then $\tilde{\mathbb{P}}$ is a probability measure. Furthermore, if X is a nonnegative random variable, then

$$\tilde{\mathbb{E}}[X] = \mathbb{E}[XZ]$$

If Z is almost surely strictly positive, we also have

$$\mathbb{E}[Y] = \tilde{\mathbb{E}}\left[\frac{Y}{Z}\right]$$

for every nonnegative random variable Y.[1]

©2016 The Trustees of the Stevens Institute of Technology

Equivalent Measure

Definition 1.6.3: Let Ω be a nonempty set and $\mathcal{F}$ a σ-algebra of subsets of Ω. Two probability measures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ are said to be equivalent if they agree which sets in $\mathcal{F}$ have probability zero.[1]

Definition 1.6.5: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\tilde{\mathbb{P}}$ be another probability measure on $(\Omega, \mathcal{F})$ that is equivalent to $\mathbb{P}$, and let Z be an almost surely positive random variable that relates $\mathbb{P}$ and $\tilde{\mathbb{P}}$ via Theorem 1.6.1. Then Z is called the Radon-Nikodým derivative of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$, and we write

$$Z = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}$$

[1]

©2016 The Trustees of the Stevens Institute of Technology

Filtration

Definition 2.1.1: Let Ω be a nonempty set. Let T be a fixed positive number and assume that for each $t \in [0, T]$ there is a σ-algebra $\mathcal{F}(t)$. Assume further that if $s \leq t$, then every set in $\mathcal{F}(s)$ is also in $\mathcal{F}(t)$. Then we call the collection of σ-algebras $\mathcal{F}(t)$, $0 \leq t \leq T$, a filtration
Definition 2.1.3: Let X be a random variable defined on a nonempty sample space Ω. The σ-algebra generate by X, denoted $\sigma(X)$, is the collection of all subsets of Ω of the form $\{X \in B\}$, where B ranges over the Borel subsets of $\mathbb{R}$.[1]

©2016 The Trustees of the Stevens Institute of Technology

Definition 2.1.5: Let X be a random variable defined on a nonempty sample space Ω. Let $\mathcal{G}$ be a σ-algebra of subsets of Ω. If every set in $\sigma(X)$ is also in $\mathcal{G}$, we say that X is $\mathcal{G}$-measurable.

Definition 2.1.6: Let Ω be a nonempty sample space equipped with a filtration $\mathcal{F}(t), 0 \leq t \leq T$. Let $X(t)$ be a collection of random variable indexed by $t \in [0, T]$. We say that this collection of random variables is an adapted stochastic process if, for each t, the random variable $X(t)$ is $\mathcal{F}(t)$-measurable.[1]

©2016 The Trustees of the Stevens Institute of Technology

Independence

Definition 2.2.1: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$. We say these two σ-algebras are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) * \mathbb{P}(B) \text{ for all } A \in \mathcal{G}, B \in \mathcal{H}$$

Let X and Y be random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. We say these two random variables are independent if the σ-algebras they generate, $\sigma(X)$ and $\sigma(Y)$, are independent. We say that the random variable X is independent of the σ-algebra $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.[1]

©2016 The Trustees of the Stevens Institute of Technology

Variance, Covariance, etc.

Definition 2.2.9: Let X be a random variable whose expected value is defined. The variance of X, denoted $Var(X) = \mathbb{V}(X)$, is

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

Because $(X - \mathbb{E}[X])^2$ is nonnegative, $\mathbb{V}(X)$ is always defined, although it may be infinite. The standard deviation of X is $\sqrt{\mathbb{V}(X)}$. The linearity of expectations shows that

$$\mathbb{V}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

©2016 The Trustees of the Stevens Institute of Technology

Let Y be another random variable and assume that $\mathbb{E}[X], \mathbb{V}(X), \mathbb{E}[Y], \mathbb{V}(Y)$ are all finite. The covariance of X and Y is

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

The linearity of expectations shows that

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

In particular, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ if and only if $Cov(X, Y) = 0$. Assume, in addition to the finiteness of expectations and variances, that $\mathbb{V}(X) > 0$ and $\mathbb{V}(Y) > 0$. The correlation coefficient of X and Y is

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

If $\rho(X, Y) = 0$, we say that X and Y are uncorrelated.[1]

©2016 The Trustees of the Stevens Institute of Technology

Definition 2.3.1: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let X be a random variable that is either nonnegative or integrable. The **conditional expectation of X given $\mathcal{G}$**, denoted $\mathbb{E}[X|\mathcal{G}]$, is any random variable that satisfies

1. **Measurability:** $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable, and
2. **Partial Averaging:**

$$\int_A \mathbb{E}[X|\mathcal{G}] d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega), \text{ for all } A \in \mathcal{G}$$

If $\mathcal{G}$ is the σ-algebra generated by some other random variable W (i.e. $\mathcal{G} = \sigma(W)$), we generally write $\mathbb{E}[X|W]$ rather than $\mathbb{E}[X|\sigma(W)]$.[1]

©2016 The Trustees of the Stevens Institute of Technology

Theorem 2.3.2: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.

1. (Linearity of conditional expectations)
$$\mathbb{E}[c_1 X + c_2 Y|\mathcal{G}] = c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}]$$
2. (Taking out what is known) if X is $\mathcal{G}$-measurable
$$\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$$
3. (Iterated conditioning) If $\mathcal{H} \subset \mathcal{G}$
$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$$
4. (Independence) If X is independent of $\mathcal{G}$ then
$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$$
5. (Conditional Jensen's Inequality) If $\varphi(x)$ is a convex function then
$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbb{E}[X|\mathcal{G}])$$

[1]

©2016 The Trustees of the Stevens Institute of Technology

Martingale

Definition 2.3.5: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, Let T be a fixed positive number, and let $\mathcal{F}(t), 0 \le t \le T$, be a filtration of sub-σ-algebras of $\mathcal{F}$. Consider an adapted stochastic process $M(t), 0 \le t \le T$.

1. If
$$\mathbb{E}[M(t)|\mathcal{F}(s)] = M(s), \text{ for all } 0 \le s \le t \le T,$$
we say this process is a martingale.
2. If
$$\mathbb{E}[M(t)|\mathcal{F}(s)] \ge M(s), \text{ for all } 0 \le s \le t \le T,$$
we say this process is a submartingale
3. If
$$\mathbb{E}[M(t)|\mathcal{F}(s)] \le M(s), \text{ for all } 0 \le s \le t \le T,$$
we say this process is a supermartingale

[1]

©2016 The Trustees of the Stevens Institute of Technology

Markov

Definition 2.3.6: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let T be a fixed positive number, and let $\mathcal{F}(t), 0 \leq t \leq T$, be a filtration of sub-σ-algebras of $\mathcal{F}$. Consider an adapted stochastic process $X(t), 0 \leq t \leq T$. Assume that for all $0 \leq s \leq t \leq T$ and for every nonnegative, Borel-measurable function f, there is another Borel-measurable function g such that

$$\mathbb{E}[f(X(t))|\mathcal{F}(s)] = g(X(s))$$

Then we say that the X is a Markov process[1]

©2016 The Trustees of the Stevens Institute of Technology

[1] S.E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Number v. 11.

©2016 The Trustees of the Stevens Institute of Technology