Stochastonomicon

Jyotirmoy Bhattacharya

May 26, 2016

Contents

Preface **v**

1 Weak Convergence **1**
1.1 Weak Convergence 1
1.2 Unbounded functions 1

2 The Multivariate Normal Distribution **3**
2.1 Conditional distribution 3

A Metric Spaces **5**
A.1 The Distance Function 5
A.2 Urysohn's Theorem 5
A.3 Partitions of Unity 6
A.4 The space $C(X)$ 7

Preface

You should not read this. Here's why:

1. This is work in progress and is likely to contain many errors.
2. There is nothing original here. Everything in these notes I found in a book or on the Internet and at best rearranged.
3. Thinking through stuff on your own is likely to be more productive.

But if you want to see a poor fellow trudge his way painfully through the landscape of modern probability, welcome!

You can download the latest version of this document from `https://www.jyotirmoy.net/misc/probability-guide.pdf`.

Please send comments and corrections to jyotirmoy@jyotirmoy.net.

Chapter 1

Weak Convergence

1.1 Weak Convergence

If we say that a sequence of measures μ_n converges to a measure μ, what should we require? The following come to mind:

1. The measure of each measurable set should converge uniformly.
2. The measure of each measurable set should converge.
3. (If the state space is a topological space) The expectation of each continuous function of the state should converge.
4. (If the state space is a topological space) The expectation of each continuous bounded function of the state should converge.

In the classical central limit theorem setting we can only prove weak convergence [TODO: Give example]

In the Markov chain setting we can do better and provide TV convergence (give citation).

For economic problems weak convergence should be enough? [TODO: What does this even mean?]

1.2 Unbounded functions

If $\mu_n \Rightarrow \mu$ then can we say that even for unbounded functions g we must have

$$\int g\, d\mu_n \to \int g\, d\mu$$

No. Take $\mu_n = (1 - 1/n)\delta_0 + (1/n)\delta_n$ and $g(x) = x$.

Chapter 2

The Multivariate Normal Distribution

Jacod and Protter's *Probability Essentials* has a nice chapter on the multivariate normal distribution.

2.1 Conditional distribution

Let (x, y) be a $\mathbb{R}^{n+m}$ dimensional multivariate normal vector where x is n-dimensional and y is m-dimensional respectively. Let the mean of (x, y) be $(0, 0)$ and its variance-covariance matrix be nonsingular.

What is the conditional distribution of x given y?

The answer is easy if x and y are independent, because then the conditional distribution of x is just the unconditional distribution, which is a multivariate normal distribution because $a'x = (a, 0)'(x, y)$ is normal for every choice of a.

What if x and y are not independent. Our approach will be to reduce to the easier independent case. We know that for components of a multivariate normal vector, zero covariance implies independence. We use the Gram-Schmidt trick of subtracting a suitably scaled copy of one variable from another to get zero covariance.

Define

$$w = y - Ax$$

We want $E[wx'] = 0$ which gives us

$$A = E(yx')[E(xx')]^{-1}$$

We know that $E(xx')$ is nonsingular since if it were not there would be a non-zero a such that $E(xx')a = 0$, so that $E[(a'x)(a'x)'] = 0$, so that $(a, 0)'E[(x, y)(x, y)'](a, 0) = 0$, which is inconsistent with $E[(x, y)(x, y)']$ being nonsingular as assumed.

Since (x, w) is obtained as a linear transformation of (x, y) the former is also a multivariate normal vector. Moreover, x and w have zero covariance by construction and hence are independent. So the conditional distribution of w given x is just the unconditional distribution of x which is just a normal.

Now $y = w + Ax$, so conditional on x, y is just w shifted by a constant, so if w has a conditional normal distribution, so has y.

To pin down the exact distribution of w we compute its mean and variance-covariance matrix.

For the mean, since w and x are independent

$$\begin{aligned}\mu_{y|x} &= E(y \mid x) \\ &= E(w \mid x) + Ax\end{aligned}$$

since w and x are independent,

$$= E(w) + Ax$$

since $E(x) = E(y) = 0$

$$\begin{aligned}&= 0 + Ax \\ &= E(yx')[E(xx')]^{-1}x\end{aligned}$$

For the conditional variance covariance

$$\begin{aligned}\Sigma_{y|x} &= E\{[y - E(y \mid x)]'[y - E(y \mid x)] \mid x\} \\ &= E\{(y - Ax)'(y - Ax) \mid x\} \\ &= E(ww' \mid x)\end{aligned}$$

since w is independent of x

$$\begin{aligned}&= E[ww'] \\ &= E[(y - Ax)(y - Ax)'] \\ &= E(yy') - AE(xy') - E(yx')A + AE(xx')A\end{aligned}$$

substituting the value of A

$$= E(yy') - E(yx')[E(xx')]^{-1}E(xy')$$

Appendix A

Metric Spaces

A.1 The Distance Function

Where not otherwise specified X is a metric space with metric $d(\cdot,\cdot)$.

Definition A.1. *If F is a non-empty closed set and $x \in X$ we define*

$$d(x, F) = \inf_{y \in F} d(x, y)$$

.

Proposition A.2. *For a fixed F, $d(x, F)$ is an uniformly continuous function of F. Indeed,*

$$|d(x, F) - d(y, F)| \leq d(x, y)$$

Proposition A.3. *If F is a non-empty closed set and $x \notin F$ then*

$$d(x, F) > 0$$

Definition A.4. *If F_1 and F_2 are non-empty disjoint closed sets, we define*

$$d(F_1, F_2) = \inf_{x \in F_1} d(x, F_2)$$

Proposition A.5. *Under conditions of Definition A.4, $d(F_1, F_2) > 0$.*

A.2 Urysohn's Theorem

Theorem A.6 (Urysohn)**.** *If F_1 and F_2 are disjoint closed sets then there exists a continuous function f such that $f(x) = 0$ for $x \in F_1$ and $f(x) = 1$ for $x \in F_2$.*

Proof. The following function has these properties

$$f(x) = \frac{d(x, F_1)}{d(x, F_1) + d(x, F_2)}.$$

□

Proposition A.7. *If F is a closed set and U is an open set such that $F \subset U$ then it is possible to find an open set V such that*

$$F \subset V \subset \bar{V} \subset U$$

Proof. If F is empty we can take $V = \emptyset$. If $U = X$ we can take $V = X$.

Otherwise, F and U^c are non-empty, disjoint closed sets. Let $f(x)$ be the continuous function provided by Theorem A.6 which is 0 on F and 1 on U^c and take

$$V = \{x \mid f(x) < 1/2\}.$$

Then $F \subset V$ and by the continuity of f, V is an open set. $\bar{V} \subset \{x \mid f(x) \leq 1/2\}$ and the latter is disjoint from U^c so that $\bar{V} \subset U$. □

A.3 Partitions of Unity

Theorem A.8. *Let U_i, $i = 1, \ldots, N$ be an open cover of X. We can find continuous functions f_i, $i = 1, \ldots, N$ such that $0 \leq f(x) \leq 1$, $\operatorname{supp} f_i \subset U_i$ and*

$$\sum_i f_i(x) = 1, \quad \text{for all } x \in X$$

In this case $\{f_i\}$ is called a partition of unity subordinate to $\{U_i\}$.

If we did not want $\operatorname{supp} f_i \subset U_i$ and could instead have done just with $f_i(x) = 0$ for $x \notin U_i$, it would have been enough to take

$$f_i(x) = \frac{d(x, U_i^c)}{\sum_{j=1}^N d(x, U_j^c)}.$$

But the stronger conclusion in the theorem as stated above needs some more work. The following argument is taken from lecture notes by Marius Crainic.

We first proof the following following

Proposition A.9 (Refinement)**.** *If $\{U_i\}$ is a finite open cover of X, then there is another open cover $\{V_i\}$ of X such that $\bar{V}_i \subset U_i$ for all i.*

Proof. Consider $F = X \setminus \bigcup_{j \neq 1} U_j$. Then F is closed, and it is a subset of U_1 (since the U_i cover X). By Proposition A.7 we can find an open V_1 such that $F \subset V_1 \subset \bar{V}_1 \subset U_1$.

From the definition of F and the fact that $F \subset V_1$ it follows that $\{V_1, U_2, \ldots, U_n\}$ is also an open cover of X. We now repeat the argument in the para above with this new open cover to replace U_2 with V_2 and so on. □

We are now in a position to prove the main theorem.

Proof of Theorem A.8. Let $\{U_i\}$ be the original open cover. We apply Proposition A.9 to create a refinement $\{V_i\}$ and then a further refinement $\{W_i\}$ such that $\bar{W}_i \subset V_i \subset \bar{V}_i \subset U_i$. Taking $\bar{W}_i$ and V_i^c as the two sets in Theorem A.6 we can find a continuous function g_i such that $g_i(x) = 1$ on $\bar{W}_i$ and $g_i(x) = 0$ on V_i^c. It follows that $\operatorname{supp} g_i \subset \bar{V}_i \subset U_i$. Now $\sum_{i=1}^N g_i(x) \neq 0$ for all x since W_i is a cover of X and $g_i(x) = 1$ on W_i. So defining

$$f_i(x) = \frac{g_i(x)}{\sum_{j=1}^N g_i(x)}$$

fulfils all the requirements of the theorem. □

Note: For a compact K, the standard topological theorem for LCH spaces gives a partition of unity where the functions have compact supports.

A.4 The space $C(X)$

We use $C(X)$ to denote the space of *bounded*, real-valued, continuous functions on X.

Theorem A.10. *The space $C(X)$ is separable if and only if X is compact.*

If. Adapted from Conway's *Functional Analysis.*

Suppose X is compact. For every n, X can be covered by a finite set of balls of radius $1/n$. Denote these by U_n^k for $k = 1, \ldots, K_n$. Let $\{f_n^k\}$ be a partition of unity subordinate to $\{U_n^k\}$ for a given n, which exists by Theorem A.8. Now consider $\bigcup_n \{f_n^k \mid k = 1, \ldots, K_n\}$. This is a countable set. Let $\mathcal{F}$ be the set of rational finite linear combinations of functions from the set $\bigcup_n \{f_n^k \mid k = 1, \ldots, K_n\}$. Then $\mathcal{F}$ is a countable set of continuous functions. We will show that it is dense in $C(X)$.

Consider some $\phi(x) \in C(X)$. Since X is compact, f is uniformly continuous. Take $\epsilon > 0$ as given. We can find a n such that $d(x, y) < 2/n$ implies $|\phi(x) - \phi(y)| < \epsilon$.

Pick an arbitrary x_k in U_n^k for every $k = 1, \ldots, K_n$ for the choice of n above. For each x_k pick a rational number q_k such that $|q_k - \phi(x_k)| < \epsilon$. Define

$$g(x) = \sum_{k=1}^{K_n} q_k f_n^k(x)$$

which is an element of $\mathcal{F}$.

Since the f_n^k sum up to 1, we can write

$$\phi(x) = \sum_{k=1}^{K_n} \phi(x) f_n^k(x),$$

so that we have

$$|\phi(x) - g(x)| \leq \sum_{k=1}^{K_n} |q_k - \phi(x)| f_n^k(x)$$

Now $f_n^k(x)$ is nonzero only on U_n^k and here

$$|q_k - \phi(x)| \leq |q_k - \phi(x_k)| + |\phi(x_k) - \phi(x)| \leq 2\epsilon$$

from which we conclude, again making use of the fact that f_n^k sum to 1, that

$$|\phi(x) - g(x)| \leq 2\epsilon.$$

Since the bound is independent of x our theorem is proved.

□

An alternative proof uses the fact that X compact means that it is separable. Let $\{x_1, x_2, \ldots\}$ be a countably dense subset of X. Define $h_n(x) = d(y, x_n)$ and consider the algebra of real polynomials in the $h_n(x)$. Apply the Stone-Weierstrass theorem to show that this algebra is dense in $C(X)$. Then approximate this algebra by the algebra of rational polynomials in $h_n(x)$ which is countable.

Only If. Again from Conway.

If X is not compact then there is a countably infinite subset $D = \{x_1, x_2, \ldots\}$ such that every subset of D is closed. For any subset A of D we can use Theorem A.6 to find an $f_A(x) \in C(X)$ which is 1 on A and 0 on $D \setminus A$. For $A \neq B$, the distance between f_A and f_B in the uniform metric is at least 1. Since there are uncountably many subsets of D there are uncountably many elements in $C(X)$ such that the distance between any pair of them is at least 1. So no countable subset of $C(X)$ can be dense.

□