STATS 310A: Theory of Probability I **Autumn 2016/17**

Lecture 7: October 17

Lecturer: Persi Diaconis *Scribes: Kenneth Tay*

7.1 Distribution Functions

7.1.1 One-dimensional case

Definition 7.1 *If μ is a probability on $\mathbb{R}$, we say that $F(x) = \mu(-\infty, x]$ is the* ***(cumulative) distribution function*** *of μ.*

Note: If we know F, we know μ (by the $\pi - \lambda$ theorem).

Properties of distribution functions:

1. F is monotonically increasing (not necessarily strict).
2. F is right continuous, i.e. $\lim\limits_{n\to\infty} F(x_n) = F(x)$ if $x_n \downarrow x$. This is because $\mu(A_n) \to \mu(A)$ if $A_n \downarrow A$.
3. $F(\infty) := \lim\limits_{x\to\infty} F(x) = 1$, $F(-\infty) := \lim\limits_{x\to-\infty} F(x) = 0$.

The converse is true too: If F is increasing, right-continuous, $F(\infty) = 1$, $F(-\infty) = 0$, then there exists a unique probability μ on $\mathbb{R}$ such that $\mu(-\infty, x] = F(x)$ for all x.

Examples of distribution functions:

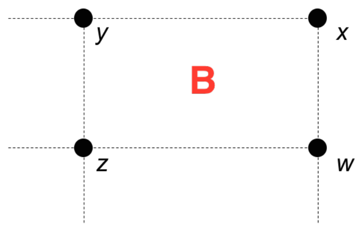
- Standard normal distribution: $F(x) = \displaystyle\int_{-\infty}^{x} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$.
- Exponential distribution: $F(x) = 1 - e^{-x}$ for $0 \le x < \infty$, $F(x) = 0$ otherwise.
- Point mass at x^*: $F(x) = 0$ for $x < x^*$, $F(x) = 1$ for $x \ge x^*$. Note that this is right-continuous but not left-continuous.

7.1.2 Multi-dimensional case

If μ is a probability on $\mathbb{R}^k$, for a set $A_x = \{y : y_i \le x_i, 1 \le x_i \le k\}$ (i.e. set of all points with all coordinates less than or equal to x), we can set $F(x) = \mu(A_x)$.

As in the 1-dimensional case, F is monotone, right-continuous and $F(\infty) = 1$, $F(-\infty) = 0$.

However, unlike the 1-dimensional case, the converse fails! For example, look at 2 dimensions. Take 4 points in the plane x, y, z, w which form the vertices of a rectangle B:



Then we have the added condition that for every x, y, z, w, we must have

$$\mu(B) = F(x) - F(y) - F(w) + F(z) \geq 0. \tag{7.1}$$

It turns out that if we have an added condition that is similar to the above, then we do have a converse. In order to make that statement precise, we need to set up a few definitions.

Definition 7.2 *In $\mathbb{R}^k$, a* ***finite rectangle*** *is a set that can be expressed as $A = \{x : a_i < x_i \leq b_i, 1 \leq i \leq k\}$.*

The class of finite rectangles is a semi-ring.

Definition 7.3 *For a finite rectangle A in $\mathbb{R}^k$, let V_A denote the set of A's vertices (there are 2^k of them). For a vertex $v \in V_A$, define*

$$sgn_A(v) = \begin{cases} -1 & \text{if } v \text{ involves an odd number of } a_i, \\ +1 & \text{if } v \text{ involves an even number of } a_i. \end{cases}$$

For a given function $F : \mathbb{R}^k \to \mathbb{R}$ and finite rectangle A in $\mathbb{R}^k$, define $\Delta_A(F) := \sum_{v \in V_A} sgn_A(v) F(v)$.

To get a feel for $\Delta_A(F)$, let's work it out for $\Delta_A(F)$ for 1 and 2 dimensions:

- For $k = 1$, let $A = (a, b]$. Then $\Delta_A(F) = F(b) - F(a)$.
- For $k = 2$, let $A = \{a_1 < x_1 \leq b_1, a_2 < x_2 \leq b_2\}$. Then $\Delta_A(F) = F(b_1, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, a_2)$, which is just Eqn 7.1.

We now state the converse theorem for $\mathbb{R}^k$:

Theorem 7.4 *Let $F(x)$ on $\mathbb{R}^k$ be a monotonically increasing, right-continuous, $F(\infty) = 1$, $F(-\infty) = 0$.*

If $\Delta_A(F) \geq 0$ for all finite rectangles A, then there exists a unique probability measure μ on the Borel sets of $\mathbb{R}^k$ such that $\mu(A) = \Delta_A(F)$ for all A.

Note: The existence of the Lebesgue measure is a special case of this theorem ($F(x) = x_1 \ldots x_n$, $\Delta_A(F) = (b_1 - a_1) \ldots (b_k - a_k)$).

Proof: Given Let $\mathcal{A}$ be the semi-ring of finite rectangles. For $A \in \mathcal{A}$, define $\mu(A) := \Delta_A(F)$. We have to prove that μ is finitely additive on $\mathcal{A}$ and countably sub-additive. Once we have done that, Theorem 11.3 from the book tells us that μ has an extension to $\sigma(\mathcal{A})$, i.e. the Borel sets of $\mathbb{R}^k$. (The extension is unique by the $\pi - \lambda$ theorem.)

We prove finite additivity in 2 steps:

Step 1: A **decomposes into a regular partition.**

That is, for every $1 \leq i \leq k$, there exist t_{ij} with $a_i = t_{i0} < t_{i1} < \cdots < t_{in_i} = b_i$, $J_{ik} = (t_{i(k-1)}, t_{ik})$, such that

$$B_{j_1 \ldots j_k} = J_{1j_1} \times \ldots \times F_{kj_k}$$

are the disjoint sets in this partition of A. Then,

$$\begin{aligned} \sum_B \mu(B) &= \sum_B \Delta_B(F) \\ &= \sum_B \sum_{v \in V_B} \operatorname{sgn}_B(v) F(v) \\ &= \sum_{v \in V_B} F(v) \sum_B \operatorname{sgn}_B(v). \end{aligned}$$

If v is an internal vertex (i.e. not a vertex of A), it is contained in at least 2 boxes. The boxes containing v can be paired off so that v has positive sign in one box and negative sign in the other. Hence, $\sum\limits_B \operatorname{sgn}_B(v) = 0$.

If v is not an internal vertex, then it is a vertex of A and $\operatorname{sgn}_B(v) = \operatorname{sgn}_A(v)$. Thus,

$$\begin{aligned} \sum_{v \in V_B} F(v) \sum_B \operatorname{sgn}_B(v) &= \sum_{v \in V_A} \operatorname{sgn}_A(v) F(v) \\ &= \mu(A). \end{aligned}$$

Step 2: A **does not decompose into a regular partition.**

Let the partition be denoted by $\mathcal{B}$. We can make regular partition $\tilde{\mathcal{B}}$ from $\mathcal{B}$ by extending the sides of the rectangles in $\mathcal{B}$. Then, using Step 1 twice, we get

$$\mu(A) = \sum_{B \in \tilde{\mathcal{B}}} \mu(B) = \sum_{B \in \mathcal{B}} \mu(B).$$

Having shown finite additivity, we can use Lemma 2(ii) of Theorem 11.4 in the book to conclude that μ is finitely sub-additive.

To show countable sub-additivity: let $A = \{x : a_i < x_i \leq b_i, 1 \leq i \leq k\} \in \mathcal{A}$, $\{A_u\}_1^\infty \in \mathcal{A}$, $A \subseteq \bigcup A_u$.

Let $A = \{x : a_i < x_i \leq b_i, 1 \leq i \leq k\}$. Define a slightly smaller rectangle $B = \{x : a_i + \delta < x_i \leq b_i, 1 \leq i \leq k\}$. Then $B \subset A$. By right-continuity, there exists a sufficiently small δ such that $\mu(A) - \varepsilon \leq \mu(B)$. Further, we can choose δ small enough so that the closure $\bar{B} \subset A$.

For each $A_u = \{x : a_i^u < x_i \leq b_i^u, 1 \leq i \leq k\}$, define a slightly larger rectangle $B_u = \{x : a_i^u < x_i \leq b_i^u + \delta, 1 \leq i \leq k\}$ such that $\mu(B_u) \leq \mu(A_u) + \varepsilon/2^u$ and the interior $B_u^\circ \supseteq A_u$.

Putting these facts together, we have

$$\bar{B} \subseteq A \subseteq \bigcup_1^\infty A_u \subseteq \bigcup_1^\infty B_u^\circ.$$

Hence, by the Heine-Borel Theorem, there exist a finite subcover of $\bar{B}$ (and B) by $B_{i_j}^{\circ}$, $1 \le j \le N$. Then

$$\begin{aligned}
\mu(A) - \varepsilon &\le \mu(B) \\
&\le \sum_{j=1}^{N} \mu(B_{i_j}) && \text{(by finite sub-additivity)} \\
&\le \sum_{j=1}^{N} \mu(A_{i_j}) + \varepsilon/2^{i_j}.
\end{aligned}$$

We get countable sub-additivity by letting ε go to zero. ∎

By setting $F(x) = \prod_{i=1}^{k} F_i(x_i)$, we obtain the following corollary:

Corollary 7.5 *If F_i's are distribution functions on $\mathbb{R}$, then $F(x) = \prod_{i=1}^{k} F_i(x_i)$ is a distribution function on $\mathbb{R}^k$.*

(In this case, $\Delta_A(F) = \prod(F_i(b_i) - F_i(a_i)) \ge 0$.)