Probability and Stochastic Processes I - Lecture 6

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/stac62/STAC622023.html

2023

II.2 Probability on Euclidean Spaces

- consider random vector $\mathbf{X} \in R^k$

note - there is an underlying (Ω, A, P) with $\mathbf{X} : \Omega \rightarrow R^k$ but this structure will only be mentioned if needed

- recall that with random vectors the basic sets we want to assign probabilities to are k-cells $(\mathbf{a}, \mathbf{b}] = X_{i=1}^k (a_i, b_i]$

note - $(\mathbf{a}, \mathbf{b}]$ can be written in terms of sets of the form $X_{i=1}^k (-\infty, b_i]$

Example II.2.1 - suppose $k = 2$

$$\begin{aligned} & (\mathbf{a}, \mathbf{b}] \\ = \; & (a_1, b_1] \times (a_2, b_2] \\ = \; & (-\infty, b_1] \times (-\infty, b_2] \backslash (-\infty, a_1] \times (-\infty, b_2] \backslash (-\infty, b_1] \times (-\infty, a_2] \end{aligned}$$

and

$$(-\infty, b_1] \times (-\infty, a_2] = (a_1, b_1] \times (-\infty, a_2] \cup (-\infty, a_1] \times (-\infty, a_2]$$

is a disjoint union so

$$\begin{aligned} P_{\mathbf{X}}((\mathbf{a}, \mathbf{b}]) = \; & P_{\mathbf{X}}((-\infty, b_1] \times (-\infty, b_2]) - P_{\mathbf{X}}((-\infty, a_1] \times (-\infty, b_2]) - \\ & P_{\mathbf{X}}((-\infty, b_1] \times (-\infty, a_2]) + P_{\mathbf{X}}((-\infty, a_1] \times (-\infty, a_2]) \end{aligned}$$

■

Definition II.2.1 For random vector $\mathbf{X} \in R^k$ the *cumulative distribution function (cdf)* $F_{\mathbf{X}} : R^k \to [0,1]$ is given by

$$F_{\mathbf{X}}(x_1, \ldots, x_k) = P_{\mathbf{X}}((-\infty, x_1] \times \cdots \times (-\infty, x_k]) = P_{\mathbf{X}}((-\infty, \mathbf{x}]). \blacksquare$$

- for any $g : R^k \to R^1$ define the i-th *difference operator* $\Delta_{a,b}^{(i)}$ by $\Delta_{a,b}^{(i)} g : R^{k-1} \to R^1$ given by

$$\begin{aligned} & (\Delta_{a,b}^{(i)} g)(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) \\ = \; & g(x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_k) - g(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_k) \end{aligned}$$

Proposition II.2.1 Any distribution function $F_{\mathbf{X}} : R^k \to [0,1]$ satisfies

(i) If $a_i \le b_i$ for $i = 1, \ldots, k$, then $P_{\mathbf{X}}((\mathbf{a}, \mathbf{b}]) = \Delta_{a_1,b_1}^{(1)} \Delta_{a_2,b_2}^{(2)} \cdots \Delta_{a_k,b_k}^{(k)} F_{\mathbf{X}}$,

(ii) $F_{\mathbf{X}}(x_1, \ldots, x_k) \downarrow 0$ as $x_i \downarrow -\infty$ and $F_{\mathbf{X}}(x_1, \ldots, x_k) \uparrow 1$ as $x_i \uparrow \infty$ for every i,

(iii) $F_{\mathbf{X}}$ is right continuous, namely, if $\delta_i \downarrow 0$ for all i, then

$$F_{\mathbf{X}}(x_1 + \delta_1, \ldots, x_k + \delta_k) \to F_{\mathbf{X}}(x_1, \ldots, x_k).$$

Proof: (i) For $k = 2$,

$$\begin{aligned}\Delta^{(1)}_{a_1,b_1}\Delta^{(2)}_{a_2,b_2}F_{\mathbf{X}} &= \Delta^{(1)}_{a_1,b_1}(F_{\mathbf{X}}(\cdot,b_2) - F_{\mathbf{X}}(\cdot,a_2)) \\ &= F_{\mathbf{X}}(b_1,b_2) - F_{\mathbf{X}}(b_1,a_2) - (F_{\mathbf{X}}(a_1,b_2) - F_{\mathbf{X}}(a_1,a_2)) \\ &= F_{\mathbf{X}}(b_1,b_2) - F_{\mathbf{X}}(b_1,a_2) - F_{\mathbf{X}}(a_1,b_2) + F_{\mathbf{X}}(a_1,a_2)\end{aligned}$$

and the result follows by Example II.2.1.

Exercise II.2.1 Show $\Delta^{(1)}_{a_1,b_1}\Delta^{(2)}_{a_2,b_2}F_{\mathbf{X}} = \Delta^{(2)}_{a_2,b_2}\Delta^{(1)}_{a_1,b_1}F_{\mathbf{X}}$ and for $k = 3$ write out $\Delta^{(1)}_{a_1,b_1}\Delta^{(2)}_{a_2,b_2}\Delta^{(3)}_{a_3,b_3}F_{\mathbf{X}}$.

(ii)

$$\begin{aligned}&\lim_{x_i\downarrow-\infty} F_{\mathbf{X}}(x_1,\ldots,x_i,\ldots,x_k) \\ = &\lim_{x_i\downarrow-\infty} P_{\mathbf{X}}((-\infty,x_1]\times\cdots\times(-\infty,x_i]\times\cdots\times(-\infty,x_k]) = 0\end{aligned}$$

because $(-\infty,x_1]\times\cdots\times(-\infty,x_i]\times\cdots\times(-\infty,x_k]$ is a monotone decreasing sequence as $x_i\downarrow-\infty$ with intersection equal to the null set and the continuity of probability.

Exercise II.2.2 Prove the second part of (ii).

(iii)

$$\begin{aligned} & \lim_{\delta_1 \downarrow 0, \ldots, \delta_k \downarrow 0} F_{\mathbf{X}}(x_1 + \delta_1, \ldots, x_k + \delta_k) \\ = & \lim_{\delta_1 \downarrow 0, \ldots, \delta_k \downarrow 0} P_{\mathbf{X}}((-\infty, x_1 + \delta_1] \times \cdots \times (-\infty, x_k + \delta_k]) = F_{\mathbf{X}}(x_1, \ldots, x_k) \end{aligned}$$

since $(-\infty, x_1 + \delta_1] \times \cdots \times (-\infty, x_k + \delta_k]$ is a monotone decreasing sequence of sets with intersection equal to $(-\infty, x_1] \times \cdots \times (-\infty, x_k]$ and the continuity of probability. ■

Theorem II.2.1 (*Extension Theorem*) If $F : R^k \to [0,1]$ satisfies
(i) $\Delta^{(1)}_{a_1,b_1}\Delta^{(2)}_{a_2,b_2}\cdots\Delta^{(k)}_{a_k,b_k}F \geq 0$ whenever $a_i \leq b_i$ for $i = 1,\ldots,k$,
(ii) $F(x_1,\ldots,x_k) \uparrow 1$ as $x_i \uparrow \infty$ for every i and $F(x_1,\ldots,x_k) \downarrow 0$ as $x_i \downarrow -\infty$ for any i
(iii) F is right continuous,

then there exists a unique probability measure P on $\mathcal{B}^k$ such that F is the distribution function of P.

note - such an F determines a probability model $(R^k, \mathcal{B}^k, P)$ and we can define a random vector with this probability model by taking $\Omega = R$ and $\mathbf{X}(\omega) = \omega$.

- so we can present $P_{\mathbf{X}}$ by the simpler $F_{\mathbf{X}}$

Example II.2.2
- define $F : R^2 \to [0,1]$ by

$$F(x_1,x_2) = \begin{cases} 0 & x_1 < 0 \text{ or } x_2 < 0 \\ 1 - e^{-x_1} - e^{-x_2} + e^{-x_1-x_2} & x_1 \geq 0 \text{ and } x_2 \geq 0 \end{cases}$$

- as we will see this satisfies the Extension Theorem and so is a cdf ■

Exercise II.2.3 In Example II.2.2 verify that
$\Delta^{(1)}_{a_1,b_1}\Delta^{(2)}_{a_2,b_2}F = (e^{-a_1} - e^{-b_1})(e^{-a_2} - e^{-b_2})$
when $0 \le a_1 \le b_1, 0 \le a_2 \le b_2$.

Exercise II.2.4 Define P on $\mathcal{B}^2$ by

$$P(B) = \begin{cases} 0 & (1,1),(-1,-1) \notin B \\ 1/2 & (1,1) \in B, (-1,-1) \notin B \\ 1/2 & (1,1) \notin B, (-1,-1) \in B \\ 1 & (1,1),(-1,-1) \in B. \end{cases}$$

Verify that P is a probability measure and determine the cdf F.

- suppose we have $F_{\mathbf{X}}$ for $\mathbf{X} \in R^k$ when $k > 2$ but we really are only interested in the probability distribution of (X_1, X_2)?

- we can get this from $F_{\mathbf{X}}$ since

$$\begin{aligned} F_{(X_1,X_2)}(x_1, x_2) &= P(X_1 \le x_1, X_2 \le x_2) \\ &= P(X_1 \le x_1, X_2 \le x_2, X_3 < \infty, \ldots, X_k < \infty) \\ &= F_{\mathbf{X}}(x_1, x_2, \infty, \ldots, \infty) \end{aligned}$$

- similarly

$$\begin{aligned} F_{X_1}(x) &= F_{\mathbf{X}}(x_1, \infty, \infty, \ldots, \infty) \\ F_{X_2}(x_2) &= F_{\mathbf{X}}(\infty, x_2, \infty, \ldots, \infty) \end{aligned}$$

- these are called the marginal distributions of the coordinates and obviously we can obtain the marginal distribution of any subvector $(X_{i_1}, \ldots, X_{i_l})$ for $1 \le l \le k$ and $i_1 < i_2 < \cdots < i_l$

Example II.2.2 (continued)

$$
\begin{aligned}
F_{X_1}(x_1) &= \begin{cases} 0 & x_1 < 0 \\ 1 - e^{-x_1} & x_1 \geq 0 \end{cases} \\
F_{X_2}(x_2) &= \begin{cases} 0 & x_2 < 0 \\ 1 - e^{-x_2} & x_2 \geq 0 \end{cases}
\end{aligned}
$$

■

- suppose we have $(R^k, \mathcal{B}^k, P_{\mathbf{X}})$
- define $p_{\mathbf{X}} : R^k \to [0,1]$ by

$$p_{\mathbf{X}}(\mathbf{a}) = P_{\mathbf{X}}(\{\mathbf{a}\}) = \lim_{\delta_1 \downarrow 0, \ldots, \delta_k \downarrow 0} P_{\mathbf{X}}((a_1 - \delta_1, a_1] \times \cdots \times (a_k - \delta_k, a_k])$$

for $\mathbf{a} \in R^k$

Definition II.3.1 The probability model $(R^k, \mathcal{B}^k, P_{\mathbf{X}})$ is *discrete* if for any $B \in \mathcal{B}^k$,

$$P_{\mathbf{X}}(B) = \sum_{\mathbf{a} \in B} p_{\mathbf{X}}(\mathbf{a})$$

and $p_{\mathbf{X}}$ is then called the *probability function* of $\mathbf{X}$. ■

Proposition II.3.1 If $(R^k, \mathcal{B}^k, P_{\mathbf{X}})$ is a discrete probability model, then there are at most countably many points $\mathbf{a} \in R^k$ such that $p_{\mathbf{X}}(\mathbf{a}) > 0$.

Proof: Let $n > 0$ and consider the set $\{\mathbf{a} : p_{\mathbf{X}}(\mathbf{a}) > 1/n\}$. If $\#(\{\mathbf{a} : p_{\mathbf{X}}(\mathbf{a}) > 1/n\}) = \infty$ Then

$$P_{\mathbf{X}}(\{\mathbf{a} : p_{\mathbf{X}}(\mathbf{a}) > 1/n\}) = \sum_{\{\mathbf{a}: p_{\mathbf{X}}(\mathbf{a}) > 1/n\}} p_{\mathbf{X}}(\mathbf{a})$$

$$\geq \sum_{\{\mathbf{a}: p_{\mathbf{X}}(\mathbf{a}) > 1/n\}} \frac{1}{n} = \frac{\infty}{n} = \infty \notin [0, 1]$$

which is a contradiction so $\#(\{\mathbf{a} : p_{\mathbf{X}}(\mathbf{a}) > 1/n\}) < \infty$ for every n which implies that

$$\#(\{\mathbf{a} : p_{\mathbf{X}}(\mathbf{a}) > 0\}) = \#(\cup_{n=1}^{\infty} \{\mathbf{a} : p_{\mathbf{X}}(\mathbf{a}) > 1/n\})$$

which is countable. ■

Proposition II.3.2 If $p : R^k \rightarrow [0,1]$ satisfies (i) $p(\mathbf{a}) \geq 0$ for all $\mathbf{a} \in R^k$ and (ii) $\sum_{\mathbf{a}\in R^k} p(\mathbf{a}) = 1$, then p defines a probability measure on $\mathcal{B}^k$ given by

$$P(B) = \sum_{\mathbf{a}\in B} p(\mathbf{a})$$

for $B \in \mathcal{B}^k$.

Proof: Clearly $0 \leq P(B) \leq 1$ for every B and $P(R^k) = 1$. Further, if $B_1, B_2, \ldots \in \mathcal{B}^k$ are mutually disjoint, then

$$P(\cup_{n=1}^{\infty} B_n) = \sum_{\mathbf{a}\in\cup_{n=1}^{\infty} B_n} p(\mathbf{a}) = \sum_{n=1}^{\infty} \sum_{\mathbf{a}\in B_n} p(\mathbf{a}) = \sum_{n=1}^{\infty} P(B_n)$$

as required. ■

Example II.3.1 *Multinomial*$(n, p_1, \ldots, p_k)$ *distribution*

- consider a wheel divided into k sectors labelled 1 through k and sector i comprises a proportion p_i of the wheel

- the wheel is spun and the sector where a pointer rests is recorded

- provided the wheel is of uniform construction and the spinning is done without control, it is reasonable to suppose that the probability of observing sector i on a spin is p_i

- suppose now that n "independent" spins are obtained with

$$X_i = \text{ the number of times sector } i \text{ is recorded}$$

- then let $\mathbf{X} = (X_1, \ldots, X_k)'$ and it is clear that $\mathbf{X}$ is a discrete random vector with $p_{\mathbf{X}}(\mathbf{a}) > 0$ iff

$$a_i \in \{0, \ldots, n\} \text{ and } a_1 + \cdots + a_k = n \qquad (*)$$

- also, because of independence the probability of getting i on the first spin and j on the second spin is

$$\begin{aligned} & P(\text{"}i \text{ on 1st spin})P(\text{"}j \text{ on 2nd spin} \mid \text{"}i \text{ on 1st spin"}) \\ = \; & P(\text{"}i \text{ on 1st spin})P(\text{"}j \text{ on 2nd spin"}) = p_i p_j \\ = \; & P(\text{"}j \text{ on 1st spin})P(\text{"}i \text{ on 2nd spin"} \mid \text{"}j \text{ on 1st spin"}) \end{aligned}$$

- so the probability of observing a_1 spins giving 1, a_2 spins giving $2, \ldots, a_k$ spins giving k, in some specified order is $p_1^{a_1} \cdots p_k^{a_k}$

- then for an **a** satisfying (*)

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{a}) &= P(X_1 = a_1, \ldots, X_k = a_k) \\ &= (\#\text{ of sequences of length } n \text{ with } a_1 \text{ 1's}, \ldots, a_k \text{ k's}) p_1^{a_1} \cdots p_k^{a_k} \\ &= \binom{n}{a_1}\binom{n-a_1}{a_2} \cdots \binom{n-a_1-\cdots-a_{k-1}}{a_k} p_1^{a_1} \cdots p_k^{a_k} \\ &= \frac{n!}{a_1!a_2!\cdots a_k!} p_1^{a_1} \cdots p_k^{a_k} = \binom{n}{a_1\ a_2\ \ldots\ a_k} p_1^{a_1} \cdots p_k^{a_k} \end{aligned}$$

which is the multinomial$(n, p_1, \ldots, p_k)$ probability function

Exercise II.3.1 (*Multivariate hypergeometric*$(N_1, \ldots, N_k$ *distribution*))
Suppose that an urn contains N balls each labelled with a number in $\{1, \ldots, k\}$ with N_i balls labelled i so $N_1 + \cdots + N_k = N$. A subset of $n \leq N$ balls is drawn out of the urn (without replacement) in such a way that it is reasonable to assign the probability $1/\binom{N}{n}$ to each such subset. Let $X_i =$ the number of balls in the sample of n labelled i. Let $\mathbf{a} = (a_1, \ldots, a_k)'$ be a possible value for the random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ so $0 \leq a_i \leq N_i$ for $i = 1, \ldots, k$ and $a_1 + \cdots + a_k = n$.
Argue that

$$p_{\mathbf{X}}(\mathbf{a}) = \frac{\binom{N_1}{a_1}\binom{N_2}{a_2}\cdots\binom{N_k}{a_k}}{\binom{N}{n}}$$

is the relevant probability function.
When $k = 3, N_1 = 3, N_2 = 3, N_3 = 2$ and $n = 4$ what are the values of (a_1, a_2, a_3) such that $p_{\mathbf{X}}(\mathbf{a}) > 0$?