Probability and Stochastic Processes I - Lecture 7

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

2023

II.4 Absolutely Continuous Probability on Euclidean Spaces

- if for $(R^k, \mathcal{B}^k, P)$ we have $P(\{\mathbf{a}\}) = 0$ for every $\mathbf{a} \in R^k$, then this is a *continuous* probability model

Example II.4.1 - suppose $k = 1$ and $F : R^1 \to [0, 1]$ is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{cases}$$

- then (i) for $a \leq b$

$$\begin{aligned} F(b) - F(a) &= \begin{cases} 0 & b < 0 \\ b & a < 0 \leq b \leq 1 \\ b - a & 0 \leq a \leq b \leq 1 \\ 1 - a & 0 \leq a \leq 1 < b \\ 1 & a \leq 0 \leq 1 < b \\ 0 & 1 \leq a \end{cases} \\ &\geq 0 \end{aligned}$$

- also (ii) $\lim_{x \to \infty} F(x) = 1$ and (iii) since F is continuous it is always right continuous

- therefore by the Extension Thm F is a cdf and since F is continuous

$$P(\{a\}) = \lim_{\delta \downarrow 0} P((a - \delta, a]) = \lim_{\delta \downarrow 0} F(a) - F(a - \delta) = 0$$

and so F corresponds to a continuous distribution on R^1 ■

- we use the notation $f : (R^k, \mathcal{B}^k) \to (R^1, \mathcal{B}^1)$ to mean $f : R^k \to R^1$ and $f^{-1}B \in \mathcal{B}^k$ for every $B \in \mathcal{B}^1$

Definition II.4.1 A probability model $(R^k, \mathcal{B}^k, P)$ is *absolutely continuous* if there is a function $f : (R^k, \mathcal{B}^k) \to (R^1, \mathcal{B}^1)$ such that

$$P(A) = \int_A f(\mathbf{x})\, d\mathbf{x}$$

for every $A \in \mathcal{B}^k$. The function f is called the *probability density function* (pdf) of the model. ■

note - $P(\{\mathbf{a}\}) = \int_{\{\mathbf{a}\}} f(\mathbf{x})\, d\mathbf{x} = 0$ and so an absolutely continuous (a.c.) model is a continuous model but there are continuous models that are not absolutely continuous

- why is f called a density?

- let $B_\delta(\mathbf{a}) = \{\mathbf{x} : ||\mathbf{x} - \mathbf{a}|| \leq \delta\}$ = ball of radius δ centered at $\mathbf{a}$

if f is continuous at $\mathbf{a}$ then (fact)

$$f(\mathbf{a}) = \lim_{\delta \downarrow 0} \frac{1}{Vol(B_\delta(\mathbf{a}))} \int_{B_\delta(\mathbf{a})} f(\mathbf{x})\, d\mathbf{x} = \lim_{\delta \downarrow 0} \frac{P(B_\delta(\mathbf{a}))}{Vol(B_\delta(\mathbf{a}))}$$

so for small δ

$$f(\mathbf{a}) \approx \frac{P(B_\delta(\mathbf{a}))}{Vol(B_\delta(\mathbf{a}))}$$

- so $f(\mathbf{a})$ is approximately the amount of probability per unit volume at $\mathbf{a}$

Proposition II.4.1 For a.c. model $(R^k, \mathcal{B}^k, P)$ with density f

(i) $f(\mathbf{x}) \geq 0$ with probability 1,

(ii) $\int_{R^k} f(\mathbf{x})\, d\mathbf{x} = 1$,

(iii) $F(x_1, \ldots, x_k) = \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(z_1, \ldots, z_k)\, dz_1 dz_2 \cdots dz_k$,

(iv) when f is continuous at $(x_1, \ldots, x_k)$, then $f(x_1, \ldots, x_k) = \frac{\partial^k F(x_1, \ldots, x_k)}{\partial x_1 \cdots \partial x_k}$.

Proof (i) $0 \leq P(f^{-1}(-\infty, 0)) = \int_{f^{-1}(-\infty,0)} f(\mathbf{x})\, d\mathbf{x} \leq 0$ and so $P(f^{-1}(-\infty, 0)) = 0$ which implies $P(f^{-1}[0, \infty)) = 1$.

(ii) $1 = P(R^k) = \int_{R^k} f(\mathbf{x})\, d\mathbf{x}$.

(iii)

$$\begin{aligned} F(\mathbf{x}) &= P((-\infty, x_1] \times \cdots \times (-\infty, x_k]) = \int_{(-\infty,x_1] \times \cdots \times (-\infty,x_k]} f(\mathbf{z})\, d\mathbf{z} \\ &= \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(z_1, \ldots, z_k)\, dz_1 dz_2 \cdots dz_k \end{aligned}$$

(iv)

$$
\begin{aligned}
& \frac{\partial F(x_1,\ldots,x_k)}{\partial x_k} \\
= \; & \frac{\partial}{\partial x_k}\int_{-\infty}^{x_k}\left(\int_{-\infty}^{x_{k-1}}\cdots\int_{-\infty}^{x_1} f(z_1,\ldots,z_k)\,dz_1\cdots dz_{k-1}\right)\,dz_k \\
= \; & \frac{\partial}{\partial x_k}\int_{-\infty}^{x_k} g(x_1,\ldots,x_{k-1},z_k)\,dz_k \\
= \; & g(x_1,\ldots,x_{k-1},x_k) \text{ by the Fundamental Thm of Calculus} \\
= \; & \int_{-\infty}^{x_{k-1}}\cdots\int_{-\infty}^{x_1} f(z_1,\ldots,z_{k-1},x_k)\,dz_1\cdots dz_{k-1}
\end{aligned}
$$

and similarly for the remaining derivatives. ■

Proposition II.4.2 If $f : (R^k, \mathcal{B}^k) \rightarrow (R^1, \mathcal{B}^1)$ satisfies (i) $f(\mathbf{x}) \geq 0$ for all $\mathbf{x}$ and (ii) $\int_{R^k} f(\mathbf{x})\, d\mathbf{x} = 1$, then f is a density for a.c. prob. model $(R^k, \mathcal{B}^k, P)$.

Proof: Consider the assignment $P(B) = \int_B f(\mathbf{x})\, d\mathbf{x}$ for $B \in \mathcal{B}^k$. Clearly $0 \leq P(B) \leq \int_{R^k} f(\mathbf{x})\, d\mathbf{x} = 1$ and so $P : \mathcal{B}^k \rightarrow [0, 1]$ and $P(R^k) = 1$. Now suppose $B_1, B_2, \ldots \in \mathcal{B}^k$ are mutually disjoint, then

$$P(\cup_{i=1}^{\infty} B_i) = \int_{\cup_{i=1}^{\infty} B_i} f(\mathbf{x})\, d\mathbf{x} = \sum_{i=1}^{\infty} \int_{B_i} f(\mathbf{x})\, d\mathbf{x} = \sum_{i=1}^{\infty} P(B_i)$$

where we have used (fact) the countable additivity of $\int$ when integrating nonnegative functions (discussed later).

- this result gives us a simple way to define a.c. probability models

Example II.4.2 *Standard Multivariate Normal Distribution on R^k*

- we write $\mathbf{X} \sim N_k(\mathbf{0}, I)$ for $\mathbf{X} \in R^k$ where $\mathbf{0} = (0, \ldots, 0) \in R^k$, $I \in R^{k \times k}$ the identity matrix when

$$\begin{aligned} f(\mathbf{x}) &= (2\pi)^{-k/2} \exp(-\mathbf{x}'\mathbf{x}/2) \\ &= (2\pi)^{-k/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} x_i^2\right) \\ &= \prod_{i=1}^{k} (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x_i^2\right) = \prod_{i=1}^{k} \varphi(x_i) \end{aligned}$$

note - when $k = 1$ then $f(\mathbf{x})$ is the $N(0,1)$ density and otherwise it is the product of $N(0,1)$ densities φ

- therefore (i) $f(\mathbf{x}) \geq 0$ for all $\mathbf{x}$ and (ii) using $\int_{-\infty}^{\infty} \varphi(x)\, dx = 1$

$$\int_{R^k} f(\mathbf{x})\, d\mathbf{x} = \int_{R^k} \prod_{i=1}^{k} \varphi(x_i)\, dx_1 \cdots dx_k = \prod_{i=1}^{k} \int_{-\infty}^{\infty} \varphi(x_i)\, dx_i = 1$$

and so by Prop II.4.2 this defines an a.c. prob. model on R^k ■

Exercise II.4.1 (a) Suppose $f(x_1, x_2, x_3) = cx_1x_2x_3$ when $(x_1, x_2, x_3) \in [0, 1]^3$ and is 0 otherwise. Determine c so that f is a pdf. (b) Using f in (a) calculate $P([1/2, 3/4] \times [2/3, 1] \times [0, 1/2])$. (c) Now suppose $f(x_1, x_2, x_3) = cx_1x_2x_3$ for $0 \leq x_1 \leq x_2 \leq x_3 \leq 1$ and is 0 otherwise. Repeat parts (a) and (b).

Exercise II.4.2 E&R 2.4.20

Exercise II.4.3 E&R 2.4.21

Exercise II.4.4 E&R 2.4.24 and 2.4.25

Exercise II.4.5 E&R 2.7.17 and 2.7.18

- suppose $\mathbf{X} \sim f_{\mathbf{X}}$ (random vector $\mathbf{X}$ has an a.c. distribution with density $f_{\mathbf{X}}$)
- then $\mathbf{X}$ has cdf

$$F_{\mathbf{X}}(x_1, \ldots, x_k) = \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_{\mathbf{X}}(z_1, \ldots, z_k)\, dz_1\, dz_2 \cdots dz_k$$

- then, for example,

$$\begin{aligned} F_{(X_1,X_2)}(x_1, x_2) &= F_{\mathbf{X}}(x_1, x_2, \infty, \ldots, \infty) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_{\mathbf{X}}(z_1, \ldots, z_k)\, dz_1\, dz_2 \cdots dz_k \\ &= \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(z_1, \ldots, z_k)\, dz_3 \cdots dz_k\, dz_1\, dz_2 \\ &= \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} g(z_1, z_2)\, dz_1\, dz_2 \end{aligned}$$

where

$$g(z_1, z_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(z_1, \ldots, z_k)\, dz_3 \cdots dz_k$$

- since the cdf determines all the probabilities (Extension Thm) we know now that (X_1, X_2) has an a.c. prob dist. with density

$$
\begin{aligned}
f_{(X_1,X_2)}(x_1, x_2) &= \frac{\partial^2 F_{(X_1,X_2)}(x_1, x_2)}{\partial x_1 \partial x_2} = g(x_1, x_2) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, z_3, \ldots, z_k)\, dz_3 \cdots dz_k
\end{aligned}
$$

- the general result holds: if we have a joint density $f_{\mathbf{X}}$ for $\mathbf{X}$ then any subvector $(X_{i_1}, \ldots, X_{i_1})$ has an a.c. distribution with density obtained by integrating out all the remaining variables

Example II.4.2 *Standard Multivariate Normal Distribution on* R^k (continued)

- suppose $\mathbf{X} \sim N_k(\mathbf{0}, I)$ which has density $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{k} \varphi(x_i)$

- then clearly $(X_{i_1}, \ldots, X_{i_l})'$ has density $\prod_{j=1}^{l} \varphi(x_{i_j})$ and so

$$
(X_{i_1}, \ldots, X_{i_l})' \sim N_l(\mathbf{0}, I)
$$

- in particular $X_i \sim N(0, 1)$ for $i = 1, \ldots, k$ ■

Exercise II.4.6 Suppose $\mathbf{X} \sim p_{\mathbf{X}}$ (random vector $\mathbf{X}$ has discrete distribution with prob. fn $p_{\mathbf{X}}$). Show that any subvector $(X_{i_1}, \ldots, X_{i_1})'$ has a discrete distribution and show how you would compute its probability function.

Exercise II.4.7 Suppose $\mathbf{X} \sim \text{multinomial}(n, p_1, p_2, p_3, p_4)$. Determine the probability functions of X_1 and (X_1, X_2). State a general result for the marginals of a multinomial distribution.

Exercise II.4.8 Suppose $\mathbf{X} \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_{k+1})$ as in Ex. II.4.5. Determine the density functions of X_1 and (X_1, X_2). State a general result for the marginals of a Dirichlet distribution.