Probability and Stochastic Processes I - Lecture 9

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

2023

absolutely continuous case

- suppose $\mathbf{X} \in R^k$ has density function $f_{\mathbf{X}}$ and we want the distribution of $\mathbf{Y} = T(\mathbf{X}) \in R^l$ where $l \leq k$

- as noted $\mathbf{Y}$ could have a discrete distribution but our interest here is in the situations where $\mathbf{Y}$ also has an a.c. distribution with density $f_{\mathbf{Y}}$ which we want to determine

- one approach to this (which can be carried out sometimes) is through the cdf

$$
\begin{aligned}
f_{\mathbf{Y}}(y_1, \ldots, y_k) &= \frac{\partial^k F_{\mathbf{Y}}(y_1, \ldots, y_k)}{\partial y_1 \cdots \partial y_k} \\
&= \frac{\partial^k P_{\mathbf{X}}(T^{-1}\{(-\infty, y_1] \times \cdots \times (-\infty, y_k]\})}{\partial y_1 \cdots \partial y_k}
\end{aligned}
$$

- this will generally work with projections T when there is a formula for $F_{\mathbf{X}}$

Example II.5.1 (Example II.2.2 Continued)

- we defined $F : R^2 \to [0,1]$ by

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

but we didn't actually prove it is a cdf (via the Extension Thm)

- but if it is, then

$$f(x_1,x_2) = \frac{\partial^2 F(x_1,x_2)}{\partial x_1 \partial x_2} = \begin{cases} 0 & x_1 < 0 \text{ or } x_2 < 0 \\ e^{-x_1-x_2} & x_1 \geq 0 \text{ and } x_2 \geq 0 \end{cases}$$

and we see that (i) $f(x_1,x_2) \geq 0$ for all (x_1,x_2) and (ii)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)\,dx_1\,dx_2 = \int_0^{\infty}\int_0^{\infty} e^{-x_1-x_2}\,dx_1\,dx$$

$$= \int_0^{\infty} e^{-x_1}\,dx_1 \int_0^{\infty} e^{-x_2}\,dx_2 = \left(-e^{-x_1}\Big|_0^{\infty}\right)\left(-e^{-x_2}\Big|_0^{\infty}\right) = 1$$

- so f is a valid pdf and thus F is a valid cdf since

$$F(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(z_1,z_2)\,dz_1\,dz_2$$

- therefore, if $Y = T(X_1, X_2) = X_1$, then

$$F_{X_1}(x_1) = F(x_1, \infty) = \begin{cases} 0 & x_1 < 0 \\ 1 - e^{-x_1} & x_1 \geq 0 \end{cases}$$

so

$$f_{X_1}(x_1) = \frac{\partial F_{X_1}(x_1)}{\partial x_1} = \begin{cases} 0 & x_1 < 0 \\ e^{-x_1} & x_1 \geq 0 \end{cases}$$

and similarly for X_2, namely, both X_1 and X_2 have exponential(1) distributions ■

- generally, we need alternative methods to determine $f_{\mathbf{Y}}$

Example II.5.2

- suppose $y = T(x_1, x_2) = x_1 + x_2$ and (X_1, X_2) has density

$$f(x_1, x_2) = \begin{cases} 2 & \text{if } 0 < x_1 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

so

$$\begin{aligned} F_Y(y) &= P_Y((-\infty, y]) = P_{(X_1, X_2)}(\{(x_1, x_2) : x_1 + x_2 \leq y\}) \\ &= \begin{cases} 0 & y < 0 \\ \int_0^{y/2} \int_{x_1}^{y-x_1} 2 dx_2 dx_1 = y^2/2 & 0 \leq y \leq 1 \\ 1 - \int_{y/2}^{1} \int_{y-x_2}^{x_2} 2 dx_1 dx_2 = 2y - y^2/2 - 1 & 1 \leq y \leq 2 \\ 1 & 2 < y \end{cases} \\ f_Y(y) &= \begin{cases} 0 & y \leq 0 \text{ or } y \geq 2 \\ y & 0 < y < 1 \\ 2 - y & 1 \leq y < 2 \end{cases} \end{aligned}$$

the *triangular density* ■

change of variable

- suppose now $T : R^k \to R^k$ is 1-1 and smooth (all 1st order partial derivatives exist and are continuous)

- so

$$T(\mathbf{x}) = \begin{pmatrix} T_1(\mathbf{x}) \\ \vdots \\ T_k(\mathbf{x}) \end{pmatrix}$$

and put

$$J_T(\mathbf{x}) = \left| \det \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial T_k(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_k(\mathbf{x})}{\partial x_k} \end{pmatrix} \right|^{-1}$$

- $J_T(\mathbf{x})$ indicates how T is changing volume at $\mathbf{x}$ since (fact)

$$J_T(\mathbf{x}) = \lim_{\delta \downarrow 0} \frac{\text{vol}(B_\delta(\mathbf{x}))}{\text{vol}(TB_\delta(\mathbf{x}))}$$

so $J_T(\mathbf{x}) < 1$ means T expands volume at $\mathbf{x}$ and $J_T(\mathbf{x}) > 1$ means T contracts volumes at $\mathbf{x} = T^{-1}(\mathbf{y})$

- now if $\mathbf{Y} = T(\mathbf{X})$, then for small δ

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &\approx \frac{P_{\mathbf{Y}}(TB_\delta(T^{-1}(\mathbf{y})))}{\text{vol}(TB_\delta(T^{-1}(\mathbf{y})))} = \frac{P_{\mathbf{X}}(B_\delta(T^{-1}(\mathbf{y})))}{\text{vol}(B_\delta(T^{-1}(\mathbf{y}))} \frac{\text{vol}(B_\delta(T^{-1}(\mathbf{y})))}{\text{vol}(TB_\delta(T^{-1}(\mathbf{y})))} \\
&\approx f_{\mathbf{X}}(T^{-1}(\mathbf{y}))J_T(T^{-1}(\mathbf{y}))
\end{aligned}
$$

- this intuitive argument can be made rigorous to prove the following

Proposition II.5.1 (*Change of Variable*) When $T : R^k \to R^k$ is 1-1, smooth and $\mathbf{Y} = T(\mathbf{X})$ where $\mathbf{X}$ has an a.c. distribution with density $f_{\mathbf{X}}$, then $\mathbf{Y}$ has an a.c. distribution with density

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(T^{-1}(\mathbf{y}))J_T(T^{-1}(\mathbf{y})).$$

Example II.5.3

- $f(x) = 1/2$ for $0 < x < 2$ (the Uniform$(0, 2)$ distribution)
- let $y = T(x) = x^2$ so $T^{-1}(y) = y^{1/2}$ and $J_T(x) = |\det(2x)|^{-1} = 1/2x$ for $x \in (0, 2)$
- note T contracts lengths on $(0, 1/2)$ and expands lengths on $(1/2, 2)$
- then

$$\begin{aligned} f_Y(y) &= f(T^{-1}(y))J_T(T^{-1}(y)) \\ &= f(y^{1/2})\frac{1}{2y^{1/2}} \\ &= \begin{cases} 0 & y \leq 0 \text{ or } y \geq 4 \\ 1/4y^{1/2} & 0 < y < 4 \end{cases} \end{aligned}$$

■

Example II.5.4 *Prove* $\int_{-\infty}^{\infty} \varphi(x)\, dx = 1$ *for* $N(0,1)$ *pdf* φ.

- consider

$$\left(\int_{-\infty}^{\infty} \varphi(x)\, dx\right)^2 = \int_{-\infty}^{\infty} \varphi(x)\, dx \int_{-\infty}^{\infty} \varphi(y)\, dy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) dxdy$$

- make the polar coordinate change of variable $T(x,y) = (r,\theta)$ where for $r \in (0,\infty), \theta \in [0, 2\pi)$

$$(x,y) = T^{-1}(r,\theta) = (r\cos\theta, r\sin\theta)$$

- **fact** - $J_T(\mathbf{x}) = 1/J_{T^{-1}}(T(\mathbf{x}))$ in general so

$$J_{T^{-1}}(r,\theta) = \left|\det\begin{pmatrix} \frac{\partial r\cos\theta}{\partial r} & \frac{\partial r\cos\theta}{\partial\theta} \\ \frac{\partial r\sin\theta}{\partial r} & \frac{\partial r\sin\theta}{\partial\theta}\end{pmatrix}\right|^{-1}$$
$$= \left|\det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta\end{pmatrix}\right|^{-1}$$
$$= \left|r(\cos^2\theta + \sin^2\theta)\right|^{-1} = 1/r$$

- then, using $r^2 = x^2 + y^2$,

$$\left(\int_{-\infty}^{\infty} \varphi(x)\,dx\right)^2 = \int_0^{\infty}\int_0^{2\pi} \frac{r}{2\pi}\exp\left(-r^2/2\right)\,d\theta dr$$
$$= \int_0^{\infty} r\exp\left(-r^2/2\right)dr = -\exp\left(-r^2/2\right)\Big|_0^{\infty} = 1$$

- this proves $\int_{-\infty}^{\infty} \varphi(x)\,dx = 1$

■

Example II.5.5 *Affine transformations*

- consider a general *affine* transformation $T : R^k \to R^k$ given by

$$T(\mathbf{x}) = A\mathbf{x} + \mathbf{b} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1k}x_k + b_1 \\ a_{21}x_1 + \cdots + a_{2k}x_k + b_2 \\ \vdots \\ a_{k1}x_1 + \cdots + a_{kk}x_k + b_k \end{pmatrix}$$

where $\mathbf{b} \in R^k, A \in R^{k \times k}$

note $T(\mathbf{x}_1) = T(\mathbf{x}_2)$ iff $A(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$ so T is 1-1 iff A is a nonsingular (invertible) matrix and in that case $T^{-1}(\mathbf{y}) = A^{-1}(\mathbf{y} - \mathbf{b}) = \mathbf{x}$

$$J_T(\mathbf{x}) = \left| \det \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial T_k(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_k(\mathbf{x})}{\partial x_k} \end{pmatrix} \right|^{-1} = |\det A|^{-1}$$

- so if $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$ then

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(T^{-1}(\mathbf{y}))J_T(T^{-1}(\mathbf{y})) = f_{\mathbf{X}}(A^{-1}(\mathbf{y} - \mathbf{b}))\,|\det A|^{-1}$$

■

Example II.5.6 *General Multivariate Normal*

- suppose $\mathbf{Z} \sim N_k(\mathbf{0}, I)$ so $f_{\mathbf{Z}}(\mathbf{z}) = (2\pi)^{-k/2}\exp(-\mathbf{z}'\mathbf{z}/2)$ for $\mathbf{z} \in R^k$

- let $\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$ where $A \in R^{k\times k}$ is nonsingular and $\boldsymbol{\mu} \in R^k$

- then by the previous example $\mathbf{X}$ has an a.c. distribution with density

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{Z}}(A^{-1}(\mathbf{x}-\boldsymbol{\mu}))\,|\det A|^{-1} \\
&= (2\pi)^{-k/2}\exp(-((A^{-1}(\mathbf{x}-\boldsymbol{\mu}))'A^{-1}(\mathbf{x}-\boldsymbol{\mu})/2)\,|\det A|^{-1} \\
&= (2\pi)^{-k/2}\,|\det A|^{-1}\exp(-((\mathbf{x}-\boldsymbol{\mu})'(A^{-1})'A^{-1}(\mathbf{x}-\boldsymbol{\mu})/2) \\
&= (2\pi)^{-k/2}\,\left|\det A \det A'\right|^{-1/2}\exp(-((\mathbf{x}-\boldsymbol{\mu})'(AA')^{-1}(\mathbf{x}-\boldsymbol{\mu})/2) \\
&= (2\pi)^{-k/2}\,\left|\det AA'\right|^{-1/2}\exp(-((\mathbf{x}-\boldsymbol{\mu})'(AA')^{-1}(\mathbf{x}-\boldsymbol{\mu})/2) \\
&= (2\pi)^{-k/2}(\det\Sigma)^{-1/2}\exp(-((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})/2)
\end{aligned}$$

where $\Sigma = AA' \in R^{k\times k}$

- when a random vector $\mathbf{X}$ has this pdf we write $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$

note - $\Sigma' = (AA')' = (A')'A' = AA' = \Sigma$ so it is a symmetric matrix and for any vector $\mathbf{x} \in R^k$

$$\mathbf{x}'\Sigma\mathbf{x} = \mathbf{x}'AA'\mathbf{x} = (A'\mathbf{x})'A'\mathbf{x} = ||A'\mathbf{x}||^2 \geq 0$$

and $||A'\mathbf{x}||^2 = 0$ iff $A'\mathbf{x} = \mathbf{0}$ which is true iff $\mathbf{x} = \mathbf{0}$ and so Σ is a *positive definite* matrix ■

Exercise II.5.4 Suppose $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$ where $A \in R^{k \times k}$ is nonsingular and $\boldsymbol{\mu} \in R^k$. Prove that $\mathbf{Y} \sim N_k(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A')$.

Exercise II.5.5 Suppose $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$ and $\Sigma = CC'$ where $C \in R^{k \times k}$ is nonsingular. Prove that $\mathbf{Z} = C^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim N_k(\mathbf{0}, I)$.

Exercise II.5.6 When $k = 2$ write out the density

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-k/2}(\det\Sigma)^{-1/2}\exp(-((\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})/2)$$

in terms of x_1 and x_2 using $\boldsymbol{\mu} = (\mu_1, \mu_2)'$,

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

and we have used the symmetry of Σ to put $\sigma_{21} = \sigma_{12}$.

Example II.5.7 *Some Properties of the Multivariate Normal*

- consider, for $\boldsymbol{\mu} \in R^k, \Sigma \in R^{k\times k}$ p.d., is

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-k/2}(\det\Sigma)^{-1/2}\exp(-((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})/2) \qquad (*)$$

a valid density so we can say $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$?

- recall the Spectral Theorem from linear algebra which says that, for any p.d. matrix $\Sigma \in R^{k\times k}$,

$$\begin{aligned}
\Sigma &= Q\Lambda Q' = \sum_{i=1}^{k} \lambda_i \mathbf{q}_i \mathbf{q}_i' \text{ where} \\
Q &= \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_k \end{pmatrix} \in R^{k\times k} \text{ orthogonal} \\
\Lambda &= \text{diag}(\lambda_1, \ldots, \lambda_k) \text{ with } \lambda_1 \geq \cdots \geq \lambda_k > 0
\end{aligned}$$

- here

$$\Sigma \mathbf{q}_j = \sum_{i=1}^{k} \lambda_i \mathbf{q}_i \mathbf{q}_i' \mathbf{q}_j = \lambda_j \mathbf{q}_j$$

since $\mathbf{q}_i'\mathbf{q}_j = 0$ when $i \neq j$ and $\mathbf{q}_j'\mathbf{q}_j = 1$, so λ_j is an eigenvalue of Σ with eigenvector $\mathbf{q}_j$

- define $\Sigma^{1/2} = Q\Lambda^{1/2}Q'$, where $\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_k^{1/2})$, called the symmetric square root of Σ since

$$\begin{aligned}(\Sigma^{1/2})' &= Q\Lambda^{1/2}Q' \\ \Sigma^{1/2}\Sigma^{1/2} &= Q\Lambda^{1/2}Q'Q\Lambda^{1/2}Q' = Q\Lambda^{1/2}I\Lambda^{1/2}Q' \\ &= Q\Lambda^{1/2}\Lambda^{1/2}Q' = Q\Lambda Q' = \Sigma\end{aligned}$$

- if $\mathbf{Z} \sim N_k(\mathbf{0}, I)$ and $A = \Sigma^{1/2}$, then Example II.5.5 shows that $X = A\mathbf{Z} + \boldsymbol{\mu} \sim N_k(\boldsymbol{\mu}, AA')$ where $AA' = \Sigma^{1/2}\Sigma^{1/2} = \Sigma$

- therefore * defines a valid pdf on R^k whenever Σ is p.d.

- clearly the level sets of $f_{\mathbf{X}}$ are given by

$$\begin{aligned}\partial E_r(\boldsymbol{\mu}, \Sigma) &= \{\mathbf{x} : (\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = r^2\} \\ &= \text{the boundary of the ellipsoidal region} \\ &\quad \text{with center at } \boldsymbol{\mu} \text{ and principal axes} \\ &\quad \text{determined by } \Sigma \text{ and } r\end{aligned}$$

Exercise II.5.7 When Σ is p.d. with spectral decomposition $Q\Lambda Q'$, then prove $\Sigma^{-1} = Q\Lambda^{-1}Q'$.

- so putting $\mathbf{w} = Q'(\mathbf{x} - \boldsymbol{\mu})$ then $\mathbf{x} = \boldsymbol{\mu} + Q\mathbf{w}$ and

$$\begin{aligned}\partial E_r(\boldsymbol{\mu}, \Sigma) &= \{\mathbf{x} : (\mathbf{x} - \boldsymbol{\mu})' Q\Lambda^{-1}Q'(\mathbf{x} - \boldsymbol{\mu}) = r^2\} \\ &= \boldsymbol{\mu} + Q\{\mathbf{w} : \mathbf{w}'\Lambda^{-1}\mathbf{w} = r^2\} \\ &= \boldsymbol{\mu} + Q(\partial E_r(\mathbf{0}, \Lambda))\end{aligned}$$

and

$$\partial E_r(\mathbf{0}, \Lambda) = \left\{\mathbf{w} : \sum \frac{w_i^2}{r^2\lambda_i} = 1\right\}$$

which is the ellipsoid in R^k with i-th semi-principal axis along the i-th standard basis vector $\mathbf{e}_i$ of length $r\lambda_i^{1/2}$

- so $\partial E_r(\boldsymbol{\mu}, \Sigma)$ has i-th semi-principal axis is on the line $\{\boldsymbol{\mu} + c\mathbf{q}_i : c \in R^1\}$ of length $r\lambda_i^{1/2}$ ■