Bayesian Statistics III/IV (MATH3341/4031)

Michaelmas term, 2021

Handout 1: Revision in probability and random variables

Lecturer & author: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

Aim

To revise a bit, linear algebra, random variables and probabilities

Linear algebra Cholesky decomposition

Probability theory Random variables, probabilities, expected values, covariance/variance matrices, characteristic function

Reading list:

- DeGroot, M. H. (1970 or 2005; Chapters 1-5). Optimal statistical decisions (Vol. 82). John Wiley & Sons.
- Strang, G. (2006). Linear Algebra and Its Applications. Cengage Learning.

1 Linear Algebra

Fact 1. [Cholesky decomposition] Every symmetric positive definite matrix $A \in \mathbb{R}^d \times \mathbb{R}^d$ can be decomposed into a product of a unique lower triangular matrix $L \in \mathbb{R}^d \times \mathbb{R}^d$ and its transpose L^{\top} , i.e.

Appendix A

$$A = LL^T$$

Matrix L is called lower triangular factor of the Cholesky decomposition, and it is often denoted as $A^{1/2} = L$.

2 Probability

- **Definition 2.** A collection $\mathscr{F} = \{A, A_1, A_2, ...\}$ of sets $A, A_1, A_2, ...$, each of which are subsets of set Ω , is called a σ -algebra if and only if
 - 1. $\Omega \in \mathscr{F}$
 - 2. If $A \in \mathscr{F}$, then $A^{\complement} \in \mathscr{F}$
 - 3. If $A_1, A_2, ... \in \mathscr{F}$ is an infinite sequence of sets in \mathscr{F} then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$
- **Definition 3.** Probability distribution P on (Ω, \mathcal{F}) is called a non-negative function if and only if
 - 1. $P(\Omega) = 1$
 - 2. If $A, B \in \mathscr{F}$ and $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$
- 3. If $A_1, A_2, ... \in \mathscr{F}$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$ then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
- **Definition 4.** We call the triple (Ω, \mathcal{F}, P) as probability space.

3 Random variables

- **Definition 5.** A d-dimensional random variable y on a probability space (Ω, \mathscr{F}, P) is a function $y : \Omega \to \mathbb{R}^d$ such as, for any subset $A \subseteq \mathbb{R}^d$, it is $\{\omega \in \Omega : y(\omega) \in A\} \in \mathscr{F}$.
- **Definition 6.** A d-dimensional random variable y on a probability space (Ω, \mathscr{F}, P) , induces a probability $P_y(\cdot)$ such that, for all subsets $A \subseteq \mathbb{R}^d$,

$$P_{y}(y \in A) = P(\{\omega \in \Omega : y(\omega) \in A\})$$

- Essentially, it induces a probability space $(\mathbb{R}^d, \mathfrak{B}, P_y)$, with \mathfrak{B} a σ -algebra containing sub-sets of \mathbb{R}^d .
- Definition 7. The (cumulative) distribution function (CDF) of a d-dimensional random variable $y \in \mathcal{Y}$ is the function $F_y : \mathbb{R}^d \to [0,1]$ such that

$$F_y(y) := F_y(y'_1, ..., y'_d) = P_y(y \in (-\infty, y'_1] \times ... \times (-\infty, y'_d])$$

- Notation 8. As $y \sim F_y$, we will denote that the random variable y follows a distribution with distribution function F_y .
- This is because the distribution function defines the distribution of the random variable.
- **Definition 9.** The d-dimensional random variable $y: \Omega \to \mathcal{Y}$ with distribution F_y is discrete, if \mathcal{Y} is a countable set and the distribution can be described by its Probability Mass Function (PMF)

$$f_y(y') := f_y(y'_1, ..., y'_d) = P(\{\omega \in \Omega : y(\omega) = y'\})$$

Definition 10. The d-dimensional random variable $y:\Omega\to\mathcal{Y}$ with distribution F_y is absolutely continuous, if \mathcal{Y} is a uncountable set and the distribution can be described by its Probability Density Function (PDF) $f_y(y)$ such that

$$P_y(y \in A) = \underbrace{\int \cdots \int}_A f_y(y_1', ..., y_d') \mathrm{d}y_1' \cdots \mathrm{d}y_d', \text{ for any } A \subseteq \mathbb{R}^d.$$

- or briefly $P_y(A) = \int_A f_y(y') dy'$, where $dy' = \prod_{i=1}^d dy'_i$.
- Fact 11. The PDF of d-dimensional random variable $y:\Omega\to\mathcal{Y}$ with CDF F_y can be computed by the partial derivative as

$$f_y(y) = \frac{d}{dt_1 \cdots dt_d} F_y(t_1, ..., t_d) \Big|_{t_1 = y_1, \cdots t_d = y_d}$$
 if F_y is differential.

4 Transforming

Fact 12. Let $y \in \mathcal{Y}$ be a d-dimensional random variable with PDF $f_y(\cdot)$. Consider a bijective function $h: \mathcal{Y} \to \mathcal{Z}$ with z = h(y), and h^{-1} its inverse. The PDF of z is

$$f_z(z) = f_y(y) \left| \det \left(\frac{dy}{dz} \right) \right| = f_y(h^{-1}(z)) \left| \det \left(\frac{d}{dz} h^{-1}(z) \right) \right|$$

- Example 13. Let $y_1 \sim \operatorname{Ex}(1/2)$, $y_2 \sim \operatorname{U}(0,2\pi)$ and independent r.v. Let $z = (z_1,z_2)^{\top}$ with $z_1 = \sqrt{y_1} \cos(y_2)$ and $z_2 = \sqrt{y_1} \sin(y_2)$. Calculate the PDF of z, and recognize its distribution.
- Hint: The PDF of r.v. x following Exponential distribution as $x \sim \text{Ex}(\lambda)$ is $f_{\text{Ex}(\lambda)}(x) = \lambda \exp(-\lambda x) \mathbb{1}(x \geq 0)$.
- Solution. It is $y=h^{-1}\left(z\right)=\left(z_1^2+z_2^2,\tan^{-1}\left(\frac{z_1}{z_2}\right)\right)^{\top}$ because

$$\begin{cases} z_1 = \sqrt{y_1} \cos(y_2) \\ z_2 = \sqrt{y_1} \sin(y_2) \end{cases} \iff \begin{cases} z_1^2 + z_2^2 = y_1 \\ \frac{z_1}{z_2} = \tan(y_2) \end{cases} \iff \begin{cases} y_1 = z_1^2 + z_2^2 \\ y_2 = \tan^{-1} \left(\frac{z_1}{z_2}\right) \end{cases}$$

50 Yet.

$$\frac{\mathrm{d}y}{\mathrm{d}z} = \begin{bmatrix} 2z_1 & 2z_2 \\ \frac{z_2}{z_1^2 + z_2^2} & -\frac{z_1}{z_1^2 + z_2^2} \end{bmatrix} \implies \det\left(\frac{\mathrm{d}y}{\mathrm{d}z}\right) = -2 \implies \left|\det\left(\frac{\mathrm{d}y}{\mathrm{d}z}\right)\right| = 2$$

Finally, because y_1 and y_2 are independent

$$f_y(y) = f_{y_1} \left(z_1^2 + z_2^2 \right) f_{y_2} \left(\tan^{-1} \left(\frac{z_1}{z_2} \right) \right) = \left[\frac{1}{2} \exp \left(-\frac{1}{2} \left(z_1^2 + z_2^2 \right) \right) \right] \left[\frac{1}{2\pi} 1 \left(\tan^{-1} \left(\frac{z_1}{z_2} \right) \in (0, 2\pi) \right) \right]$$

54 Hence,

$$f_z(z) = f_y(y) \left| \det \left(\frac{\mathrm{d}y}{\mathrm{d}z} \right) \right| = \frac{1}{2\pi} \exp \left(-\frac{1}{2} \left(z_1^2 + z_2^2 \right) \right) = \prod_{i=1}^2 \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} z_i^2 \right)$$

that is a bivariate standard normal distribution i.e. $z \sim N_2(0, I)$.

5 Marginalizing & Integrating out

Fact 14. Let (n+d)-dimensional random variable $y \in \mathcal{Y}$ with distribution $F_y(\cdot)$. Consider a partition $y = (x, \theta)$ where $x \in \mathcal{X}$ is n-dimensional and $\theta \in \Theta$ is d-dimensional. Then

1. the marginal CDF of x results by setting θ as ∞

$$F_x(x) = \lim_{\theta \to \infty} F_y(x, \theta) \qquad \qquad = \lim_{\theta_1 \to \infty, \dots, \theta_d \to \infty} F_y(x_1, \dots, x_n, \theta_1, \dots, \theta_d)$$

2. the marginal PDF/PMF of x results by integrating out θ (the dimensions we marginalize)

$$f_x(x) = \begin{cases} \int_{\mathbb{R}^d} f_y(x,\theta) d\theta & \text{if } \theta \text{ is cont.} \\ \sum_{\forall \theta \in \mathbb{R}^d} f_y(x,\theta) & \text{if } \theta \text{ is discr.} \end{cases} = \begin{cases} \int_{\mathbb{R}^d} f_y(x_1,...x_n,\theta_1,...,\theta_d) d\theta_1...d\theta_d & \text{if } \theta \text{ is cont.} \\ \sum_{\forall \theta \in \mathbb{R}^d} f_y(x_1,...x_n,\theta_1,...,\theta_d) & \text{if } \theta \text{ is discr.} \end{cases}$$

6 Independence

Definition 15. Given a probability space $(\Omega, \mathfrak{B}, P)$, events $A, B \in \mathfrak{B}$ are independent if and only if

$$P(A \cap B) = P(A)P(B)$$

Fact 16. Let (n+m)-dimensional random variable $y \in \mathcal{Y}$. Consider a partition y = (x, z) where $x \in \mathcal{X}$ is ndimensional and $z \in \mathcal{Z}$ is m-dimensional. Let $F(\cdot)$ denote CDF, and $f(\cdot)$ denote PDF/PMF.

• The r.v. x and y are independent if and only if, for any $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Z}$,

$$P(\lbrace x \in A \rbrace \cap \lbrace z \in B \rbrace) = P(x \in A)P(z \in B)$$

• The r.v. x and y are independent if and only if

$$F(x,z) = F(x)F(z), \ \forall x \in \mathcal{X}, \ \forall z \in \mathcal{Z}$$

• The r.v. x and z are independent if and only if

$$f(x,z) = f(x)f(z), \forall x \in \mathcal{X}, \forall z \in \mathcal{Z}$$

7 Expected value

Definition 17. Expected value of the d-dimensional random variable $y \in \mathcal{Y}$ with CDF F and PDF/PMF f is the d-dimensional quantity

$$\mathbf{E}(y) = \int y \mathrm{d}F(y) = \begin{cases} \int_{y \in \mathcal{Y}} y f(y) \mathrm{d}y, & \text{if } y \text{ is cont.} \\ \\ \sum_{y \in \mathcal{Y}} y f(y), & \text{if } y \text{ is discr.} \end{cases}$$

whose ith element is

$$\left[\mathbf{E}(y)\right]_i = \begin{cases} \int y_i f_{y_i}(y_i) \mathrm{d}y_i, & \text{if } y_i \text{ is cont.} \\ \\ \sum_{\forall y_i} y_i f_{y_i}(y_i), & \text{if } y_i \text{ is discr.} \end{cases}$$

for i=1,...,d. Here $f_{y_i}(y_i)=\int_{y\in\mathcal{Y}}f(y)\mathrm{d}y_1\cdots\mathrm{d}y_{i-1}\mathrm{d}y_{i+1}\cdots\mathrm{d}y_d$ is the marginal PMF/PDF of y_i .

Fact 18. If $y \in \mathcal{Y}$ is a d-dimensional random variable with CDF F_y and PDF/PMF $f_y(\cdot)$, and $\psi : \mathcal{Y} \to \mathbb{R}^d$ is an integrate function with $\psi(\cdot) := (\psi_1(\cdot), ..., \psi_d(\cdot))$, then

$$E(\psi(y)) = \begin{cases} \int \psi(y) f_y(y) dy, & \text{if y is cont.} \\ \\ \sum_{\forall y} \psi(y) f_y(y), & \text{if y is discr.} \end{cases}$$

s with elements

$$\left[E(\psi(y))
ight]_i = egin{cases} \int \psi_i(y) f_y(y) dy, & ext{if y is cont.} \ \\ \sum_{orall y} \psi_i(y) f_y(y), & ext{if y is discr.} \end{cases}$$

Example 19. If $(q \times k)$ -dimensional random variable $y \in \mathcal{Y}$ is a matrix, then its expectation $E(y) = \int y dF(y)$ is a matrix too whose (i, j)-th element is

$$\left[\mathbf{E}(y) \right]_{i,j} = \mathbf{E}(y_{i,j}) = \begin{cases} \int y_{i,j} f_{y_{i,j}}(y_{i,j}) \mathrm{d}y_{i,j}, & \text{if } y_{i,j} \text{ is cont.} \\ \\ \sum_{\forall y_{i,j}} y_{i,j} f_{y_{i,j}}(y_{i,j}), & \text{if } y_{i,j} \text{ is discr.} \end{cases}$$

for i=1,...,n, and j=1,...,m. Here $f_{y_{i,j}}(\cdot)$ is the marginal PMF/PDF of $y_{i,j}$.

Proposition 20. The following properties are valid

1. Let fixed matrix/vectors A, c, and z = c + Ay with suitable dimensions then

$$E(z) = E(c + Ay) = c + AE(y)$$

2. Let random variables $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$, and let functions ψ_1 and ψ_2 defined on \mathcal{Z} and \mathcal{Y} , then

$$E(\psi_1(z) + \psi_2(y)) = E(\psi_1(z)) + E(\psi_2(y))$$

3. Random variables $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$ are independent if and only if

$$E(\psi_1(z)\psi_2(y)) = E(\psi_1(z))E(\psi_2(y))$$

for any functions ψ_1 and ψ_2 defined on \mathcal{Z} and \mathcal{Y} .

Proof. Given as Exercise 8 in the Exercise Sheet.

8 Covariance matrix

Definition 21. The covariance matrix between random variable $z \in \mathcal{Z} \subseteq \mathbb{R}^d$ and random variable $y \in \mathcal{Y} \subseteq \mathbb{R}^q$ is defined as the $d \times q$ matrix

$$Cov(z, y) = E((z - E(z))(y - E(y))^{\top})$$

Proposition 22. The following properties are the direct analogues of the 1D cases

- 1. $Cov(z, y) = E(zy^{\top}) E(z) (E(y))^{\top}$
- 2. $Cov(z, y) = (Cov(y, z))^{\top}$
 - 3. $Cov(c_1 + A_1z, c_2 + A_2y) = A_1Cov(z, y)A_2^{\top}$, for fixed matrices A_1, A_2 , and vectors c_1, c_2 with suitable dimensions.
 - 4. If z and y are independent random vectors then Cov(z, y) = 0

Proof. (1)-(3) result from the definition. (4) results from Prop 20, as $E(zy^{\top}) = E(z) (E(y))^{\top}$.

Proposition 23. It can be seen that

$$[Cov(z,y)]_{i,j} = Cov(z_i,y_j)$$

for all i = 1, ..., d, and j = 1, ..., q. Namely, the (i, j)-th element of the covariance matrix between vector z and y is the covariance between their elements z_i and y_j .

Definition 24. The covariance matrix of random vector $y \in \mathcal{Y} \subseteq \mathbb{R}^d$ is defined as the $d \times d$ matrix Var(y)

$$Var(y) = Cov(y, y) = E\left((y - E(y))(y - E(y))^{\top}\right)$$

Proposition 25. It can be seen that

$$[Var(y)]_{i,j} = Cov(y_i, y_j)$$
 for all $i, j = 1, ..., d$

119 and

$$[Var(y)]_{i,i} = Var(y_i)$$
 for all $i = 1, ..., d$.

Proposition 26. The following properties are the direct analogues of the 1D cases

- 1. $Var(y) = E(yy^{\top}) E(y) (E(y))^{\top}$
- 2. $Var(c + Ay) = AVar(y)A^{\top}$, for fixed matrix A, and vectors c with suitable dimensions.
 - 3. $Var(y) \ge 0$; (semi-positive definite)

25 Proof. Given as Exercise 11 in the Exercise Sheet.

9 Characteristic function

²⁷ Characteristic functions provide an alternative way to the probability function for describing a random variable.

Definition 27. The characteristic function (CF) of a d dimensional random variable X with distribution $F(\cdot)$ is

$$\varphi_x(t) = \mathbf{E}(e^{it^T x}) = \int e^{it^T x} dF(x)$$

for $t \in \mathbb{R}^d$, where $e^{it^T x} = \cos(t^T x) + i\sin(t^T x)$.

Proposition 28. Some properties of characteristic functions

- 1. $\varphi_x(t)$ exists for all $t \in \mathbb{R}^d$ and is absolutely continuous
- 2. $\varphi_x(0) = 1$ and $|\varphi_x(t)| \leq 1$ for all $t \in \mathbb{R}^d$
- 3. $\varphi_{A+Bx}(t) = e^{it^T A} \varphi_x(B^T t)$ if $A \in \mathbb{R}^d$ and $B \in \mathbb{R}^{k \times d}$ are constants
- 4. $\varphi_{x+y}(t) = \varphi_x(t)\varphi_y(t)$ if and only if x and y are independent
 - 5. if $M_x(t) = E(e^{t^T x})$ is the moment generating function, then $M_x(t) = \varphi_x(-it)$
- 137 *Proof.* Given as Exercise 12 in the exercise sheet.
- Fact 29. Two random variables have equal characteristic functions if and only if they follow the same distribution.
- AKA: CF completely determines the probability distribution of the random variable
- Fact 30. If $\varphi_x(t)$ is absolutely integratable, then x has PDF

$$f(x) = \frac{1}{(2\pi)^d} \int_{-\infty}^{+\infty} e^{-it^T x} \varphi_x(t) dt$$

Example 31. Address the following:

- 1. If $z \sim N(0,1)$ then $\varphi_z(t) = \exp(-\frac{1}{2}t^2)$
- 2. If $x \sim N(\mu, \sigma^2)$ then $\varphi_x(t) = \exp(i\mu t \frac{1}{2}t^2\sigma^2)$
- 3. If $\varphi_z(t) = \exp(-\frac{1}{2}t^2)$ then $f_{N(0,1)}(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$

Solution. It is

1. It is

$$\begin{split} \varphi_z(t) &= \mathsf{E}(e^{itz}) = \int e^{itz} \mathsf{d} F_{\mathsf{N}(0,1)}(z) = \int e^{itz} f_{\mathsf{N}(0,1)}(z) \mathsf{d}z = \int \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2 + itz) \mathsf{d}z \\ &= \int \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2 + \frac{2}{2}itz \pm \frac{1}{2}(it)^2 z) \mathsf{d}z = \int \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(z-it)^2) \mathsf{d}z \times \exp(\frac{1}{2}(it)^2) \\ &= \exp(-\frac{1}{2}t^2) \end{split}$$

- 2. It is $\varphi_x(t) = \varphi_{\mu+\sigma z}(t) = \exp(i\mu t)\varphi_z(\sigma t) = \exp(i\mu t \frac{1}{2}t^2\sigma^2)$.
- 3 Iti

$$\begin{split} f(z) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itz} \varphi_z(t) \mathrm{d}t \ = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itz) \exp(-\frac{1}{2}t^2) \mathrm{d}t \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-\frac{1}{2}t^2 \pm \frac{1}{2}(iz)^2 + \frac{2}{2}tiz) \mathrm{d}t \ = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}(t-iz)^2 - z^2\right) \mathrm{d}t \\ &= \sqrt{\frac{1}{2\pi}} \int_{-\infty}^{+\infty} \sqrt{\frac{1}{2\pi}} \exp\left(-\frac{1}{2}(t-iz)^2\right) \mathrm{d}t \exp\left(-\frac{1}{2}z^2\right) = \sqrt{\frac{1}{2\pi}} \exp\left(-\frac{1}{2}z^2\right) = f_{\mathsf{N}(0,1)}(z) \end{split}$$

Theorem 32. The distribution of a d-dimensional random variable $x \in \mathbb{R}^d$ is completely determined be the set of all 1--dimensional distributions of linear combinations $a^{\top}x$, for any $a \in \mathbb{R}^d$.

Proof. Let $y = a^{\top} x$, for any $a \in \mathbb{R}^d$. Then for any $s \in \mathbb{R}$

$$\varphi_y(s) = \mathsf{E}(e^{is^T y}) = \mathsf{E}(e^{is^T a^T x}) = \mathsf{E}(e^{i(as)^T x}) = \mathsf{E}(e^{i\tilde{t}^T x}) = \varphi_x(\tilde{t})$$

where $\tilde{t} = as$ is any d-dimensional vector.

10 Conditioning

Definition 33. Assume a probability space (Ω, \mathscr{F}, P) . For any sets $A, B \in \mathscr{F}$, the conditional probability of A given B is defined as

$$P(A|B) = \frac{P(B \cap A)}{P(B)}$$
 if $P(B) \neq 0$.

Definition 34. Let $y \in \mathcal{Y}$ be a random variable. Consider a partition $y = (x, \theta)$ with $x \in \mathcal{X}$ and $\theta \in \Theta$. The expected value of θ conditional that random variable $x \in B \subseteq \mathcal{X}$ is

$$E(\theta|x \in B) = \frac{E(\theta 1(x \in B))}{P(x \in B)}, \quad \text{if } P(x \in B) > 0.$$

- Fact 35. Let a random variable $y \in \mathcal{Y}$ with PDF/PMF $f(\cdot)$. Consider a partition $y = (x, \theta)$ with $x \in \mathcal{X}$ and $\theta \in \Theta$.
 - 1. The conditional MPF/PDF and CDF of random variable θ given the random variable x

$$f_{\theta|x}(\theta|x) = \frac{f(\theta,x)}{f(x)}, \; ; \quad F_{\theta|x}(\theta|x) = \begin{cases} \int_{-\infty}^{\theta_1} \cdots \int_{-\infty}^{\theta_d} f_{\theta|x}(\vartheta|x) d\vartheta, & \theta, \ cont. \\ \\ \sum_{\vartheta_1 = -\infty}^{\theta_1} \cdots \sum_{\vartheta_d = -\infty}^{\theta_d} f_{\theta|x}(\vartheta|x), & \theta, \ discr. \end{cases}$$

- provided that f(x) > 0.
- The low index $\cdot_{\theta|x}$ can be omitted when obvious, e.g. $f(\theta|x) := f_{\theta|x}(\theta|x)$, and $F(\theta|x) := F_{\theta|x}(\theta|x)$.
 - 2. The expected value of θ given the random variable x

$$E(\theta|x) = \int \theta dF(\theta|x) = \begin{cases} \int \theta f(\theta|x) d\theta & \text{, if θ is cont.} \\ \\ \sum_{\forall \theta} \theta f(\theta|x) & \text{, if θ is discr.} \end{cases} provided that $f(x) > 0$$$

- Example 36. Let a random variable $y \in \mathcal{Y}$ with distribution $F(\cdot)$. Consider a partition $y = (x, \theta)^{\top}$ with $x \in \mathcal{X}$ and $\theta \in \Theta$. Then
- 1. $E(\theta) = E(E(\theta|x))$
- 2. $Var(\theta) = E(Var(\theta|x)) + Var(E(\theta|x))$
- 79 Solution.
 - 1. It is

$$E(E(\theta|x)) = \int \left(\int \theta dF(\theta|x)\right) dF(x) = \int \int \theta dF(x,\theta) = \int \int \theta dF(x|\theta) dF(\theta)$$
$$= \int \theta \left(\int dF(x|\theta)\right) F(\theta) = \int \theta F(\theta) = E(\theta)$$

2. It is

$$\begin{aligned} \operatorname{Var}(\theta) &= \operatorname{E}\left(\operatorname{E}(\theta\theta^{\top})\right) - \operatorname{E}\left(\theta\right) \operatorname{E}\left(\theta\right)^{\top} \\ &= \operatorname{E}\left(\operatorname{E}(\theta\theta^{\top}|x)\right) - \operatorname{E}\left(\operatorname{E}(\theta|x)\right) \operatorname{E}\left(\operatorname{E}(\theta|x)\right)^{\top} \\ &= \operatorname{E}\left(\operatorname{E}(\theta\theta^{\top}|x)\right) - \operatorname{E}\left(\operatorname{E}(\theta|x)\operatorname{E}(\theta|x)^{\top}\right) + \operatorname{E}\left(\operatorname{E}(\theta|x)\operatorname{E}(\theta|x)^{\top}\right) - \operatorname{E}\left(\operatorname{E}(\theta|x)\right) \operatorname{E}\left(\operatorname{E}(\theta|x)\right)^{\top} \\ &= \operatorname{E}\left(\operatorname{E}(\theta\theta^{\top}|x) - \operatorname{E}(\theta|x)\operatorname{E}(\theta|x)^{\top}\right) + \operatorname{E}\left(\operatorname{E}(\theta|x)\operatorname{E}(\theta|x)^{\top} - \operatorname{E}\left(\operatorname{E}(\theta|x)\right)\operatorname{E}\left(\operatorname{E}(\theta|x)\right)^{\top}\right) \\ &= \operatorname{E}\left(\operatorname{Var}(\theta|x)\right) + \operatorname{Var}\left(\operatorname{E}(\theta|x)\right) \end{aligned}$$

8 Conditional independence

Definition 37. Given a probability space $(\Omega, \mathfrak{B}, P)$, and events $A, B, C \in \mathfrak{B}$, A and B are conditionally independent given C if and only if

$$P(A \cap B|C) = P(A|C)P(B|C)$$
, for $P(C) > 0$.

Fact 38. Let (n+m+d)-dimensional random variable $y \in \mathcal{Y} \subseteq \mathbb{R}^{n+m+d}$. Consider a partition $y = (x, z, \theta)$ where $x \in \mathcal{X} \subseteq \mathbb{R}^n$, $z \in \mathcal{Z} \subseteq \mathbb{R}^m$, and $\theta \in \Theta \subseteq \mathbb{R}^d$.

• The r.v. x and y are independent given θ if and only if

$$F(y, z|\theta) = F(y|\theta)F(z|\theta), \ \forall x \in \mathbb{R}^n, \ \forall z \in \mathbb{R}^m$$

where $F(\cdot|\theta)$ denotes the conditional CDF.

• This implies that r.v. x and z are independent given θ if and only if

$$f(y, z|\theta) = f(y|\theta)f(z|\theta)$$

where $f(\cdot|\theta)$ denotes the conditional PDF/PMF.

11 Inverting / updating

This offers a probabilistic mechanism for (1.) inversion $(B|A) \longmapsto (A|B)$, or (2.) updating $(A) \longmapsto (A|B)$.

Fact 39. [Bayesian theorem with sets] Assume a probability space (Ω, \mathcal{F}, P) . For any sets $A, B \in \mathcal{F}$, it is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad provided that P(B) \neq 0.$$

The extension of the Bayesian theorem to the random variables is not straightforward.

Proposition 40. [Bayesian theorem with random variables] Let a random variable $y \in \mathcal{Y}$. Consider a partition $y = (x, \theta)$ with $x \in \mathcal{X}$ and $\theta \in \Theta$. Then the PDF/PMF of $\theta | x$ is

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)dF(\theta)} = \begin{cases} \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta} & \text{, if θ is cont.} \\ \\ \frac{f(x|\theta)f(\theta)}{\sum_{\forall \theta}f(x|\theta)f(\theta)} & \text{, if θ is discr.} \end{cases}$$

Proof. Given as Exercise 15, in the Exercise Sheet.

12 Practice

Question 41. Try the Exercises 13, 14, 15, from the Exercise sheet.

Supplementary material

A numerical algorithm that allows the computation of the lower triangular matrix of the Cholesky factorisation is given in Fact 42.

Fact 42. [Cholesky-Banachiewicz algorithm] The lower triangular matrix L from the Cholesky decomposition of $A \in \mathbb{R}^d \times \mathbb{R}^d$ is calculated as

$$\quad \text{for } i=1,...d$$

for
$$j=1,..,d$$

$$L_{i,j} = egin{cases} \sqrt{A_{i,i} - \sum_{k=1}^{i-1} L_{i,k}^2} &, \emph{if} i = j \ rac{1}{L_{j,j}} (A_{i,j} - \sum_{k=1}^{i-1} L_{i,k} L_{j,k}) &, \emph{if} i > j \ 0 &, \emph{if} i < j \end{cases}$$

such as $A = L^{\top}L$.

Proof. Out of the scope.

So if A is $a \mathbb{R}^3 \times \mathbb{R}^3$ matrix then the computations evolve row-wise, i.e.

$$L_{1,1} o L_{2,1} o L_{2,2} o L_{3,1} o L_{3,2} o L_{3,3}$$
 etc...