Multivariate Probability

Yingzhen Li

Department of Computing
Imperial College London

@liyzhen2
yingzhen.li@imperial.ac.uk

October 28, 2022

Recap: univariate probability

Univariate probability examples:

Bernoulli:

- X takes binary values $\{0, 1\}$
- PMF: $p(X = 1) = \rho, \quad \rho \in [0, 1]$

Categorical:

- X takes values in $\{1, ..., C\}$
- PMF satisfies $\sum_{c=1}^{C} p(X = c) = 1, \quad p(X = c) \geqslant 0$

Recap: univariate probability

Univariate probability examples:

Gaussian:

- X takes continuous real number values in $\mathbb{R}$
- PDF: $p(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}(x - \mu)^2]$
- CDF: $F(x) = P(X \leqslant x) = \int_{-\infty}^{x} p(X = \alpha) d\alpha$
- Notice that $p(x) = \frac{dF(x)}{dx}$

Lectures on multivariate probability

Topic of today and next Monday: multivariate probability

- Definitions and some examples
- Joint, marginal, and conditional distributions
- Sum rule and product rule

Lots of techniques to learn and master!

- Change-of-variables rule
- Computing mean/variance/expectations

Multivariate probability

We want to work with multiple random variables $X_1, ..., X_k$

- Reuse the concepts introduced in univariate probability:
  - Sample space Ω, Event space $\mathcal{E} = 2^{\Omega}$, Probability: $\mathbb{P} : \mathcal{E} \to [0, 1]$
- $X_n : \Omega \to V_{X_n}$ maps $\omega \in \Omega$ to some integer/real value

Multivariate probability

We want to work with multiple random variables $X_1, ..., X_k$

- Reuse the concepts introduced in univariate probability:
  - Sample space Ω, Event space $\mathcal{E} = 2^{\Omega}$, Probability: $\mathbb{P} : \mathcal{E} \rightarrow [0, 1]$
- $X_n : \Omega \rightarrow V_{X_n}$ maps $\omega \in \Omega$ to some integer/real value
- We can define the support $\mathcal{A}$ in the value space of $X_1, ..., X_n$: $\mathcal{A} = \{(x_1, x_2, ..., x_N) : X_n(\omega) = x_n, \omega \in \Omega\}$
  - This means event $E \subset \Omega$ can be mapped to a measureable set $A \subset \mathcal{A}$, so $P(A) := \mathbb{P}(E) \in [0, 1]$

Multivariate probability

We want to work with multiple random variables $X_1, ..., X_k$

- Reuse the concepts introduced in univariate probability:
  - Sample space Ω, Event space $\mathcal{E} = 2^\Omega$, Probability: $\mathbb{P} : \mathcal{E} \to [0, 1]$
- $X_n : \Omega \to V_{X_n}$ maps $\omega \in \Omega$ to some integer/real value
- We can define the support $\mathcal{A}$ in the value space of $X_1, ..., X_n$: $\mathcal{A} = \{(x_1, x_2, ..., x_N) : X_n(\omega) = x_n, \omega \in \Omega\}$
  - This means event $E \subset \Omega$ can be mapped to a measureable set $A \subset \mathcal{A}$, so $P(A) := \mathbb{P}(E) \in [0, 1]$
- Multivariate PMF/PDF satisfies $p(x_1, x_2, ..., x_N) \geqslant 0$ and:

$$\text{PMF:} \quad \sum_{(x_1, x_2, ..., x_N) \in A} p(x_1, x_2, ..., x_N) = P(A), \quad \forall A \subset \mathcal{A}.$$

$$\text{PDF:} \quad \int_A p(x_1, x_2, ..., x_N) dx_1 dx_2 ... dx_N = P(A), \quad \forall A \subset \mathcal{A}.$$

Multivariate probability

Let's say you are in a zoo that has infinite number of animals:

Ω: sample space

...

$X_1 : \Omega \to \mathbb{R}$
"height of the animal in cm"

$X_1($  $) = 9.6$

$X_1($  $) = 72.8$

...

$X_2 : \Omega \to \mathbb{R}$
"weight of the animal in kg"

$X_2($  $) = 1.02$

$X_2($  $) = 17.4$

...

$X_3 : \Omega \to \mathbb{N}^+$
"fur colour of the animal"

$X_2($  $) = 1$

$X_2($  $) = 3$

...

$(1: yellow, 2: blue, 3: orange, \ldots)$

- Support: $\mathcal{A} \subset \mathbb{R} \times \mathbb{R} \times \mathbb{N}^+$
- A measurable subset in $\mathcal{A}$ can be
  $A = \{10.0 \leqslant x_1 \leqslant 50.0, 1 \leqslant x_2 \leqslant 10.0, x_3 \in \{2, 3, 4\}\}$
  ("The animal's height, weight and fur colour are within some values/regimes")

Multivariate probability

Let's say you are in a zoo that has infinite number of animals:

Ω: sample space

...

$X_1 : \Omega \to \mathbb{R}$
"height of the animal in cm"

$X_1($  $) = 9.6$

$X_1($  $) = 72.8$

...

$X_2 : \Omega \to \mathbb{R}$
"weight of the animal in kg"

$X_2($  $) = 1.02$

$X_2($  $) = 17.4$

...

$X_3 : \Omega \to \mathbb{N}^+$
"fur colour of the animal"

$X_2($  $) = 1$

$X_2($  $) = 3$

...

$(1: yellow, 2: blue, 3: orange, \ldots)$

- Let's assume the event space $\mathcal{E} = 2^\Omega$
- Figuring out $P(A)$: find the biggest set $E \subset \Omega$ such that $(X_1, ..., X_N)(E) := \{(X_1(\omega), ..., X_N(\omega)) : \omega \in E\} \subset A$, then set $P(A) := \mathbb{P}(E)$

Multivariate probability

Let's say you are in a zoo that has infinite number of animals:

Ω: sample space

...

$X_1: \Omega \to \mathbb{R}$
"height of the animal in cm"

$X_1($  $) = 9.6$

$X_1($  $) = 72.8$

...

$X_2: \Omega \to \mathbb{R}$
"weight of the animal in kg"

$X_2($  $) = 1.02$

$X_2($  $) = 17.4$

...

$X_3: \Omega \to \mathbb{N}^+$
"fur colour of the animal"

$X_2($  $) = 1$

$X_2($  $) = 3$

...

$(1: yellow, 2: blue, 3: orange, \ldots)$

- $p(x_1, x_2, x_3)$ satisfies:

$$\int \sum_{(x_1,x_2,x_3)\in A} p(x_1,x_2,x_3)dx_1dx_2 = P(A), \quad \forall A \subset \mathcal{A}.$$

Example: multinomial distribution

Rolling a k-sided dice independently for n times, define $X_i = \#$ side i

$$k = 6, n = 6, \quad p(\text{⚀}) = p_1, \ldots \quad \textstyle\sum_{i=1}^{k} p_i = 1$$

$$\omega = (\text{⚃}, \text{⚅}, \text{⚃}, \text{⚀}, \text{⚂}, \text{⚂})$$

$$X_1 = 1, X_2 = 0, X_3 = 2, X_4 = 2, X_5 = 0, X_6 = 1$$

- Support: $\mathcal{A} = \{x_1, \ldots, x_n \in \mathbb{N} : \sum_{i=1}^{k} x_i = n\} \subset \mathbb{N}^k$
- PMF: note that permuting elements in ω does not change X_i

$$p(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

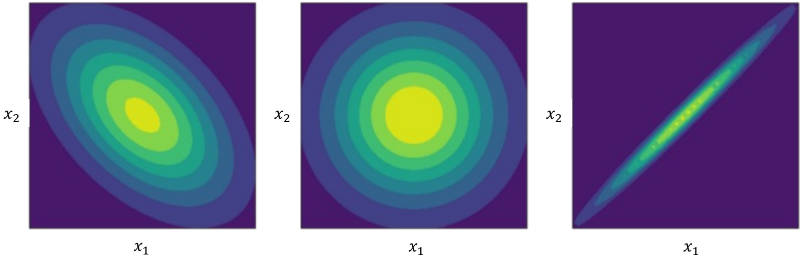
Example: multivariate Gaussian distribution

Univariate Gaussian: $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}(x-\mu)^2]$

Multivariate Gaussian: $\boldsymbol{x} = (x_1, ..., x_d)^\top$

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})]$$

- Independent Gaussians: $\Sigma = diag(\sigma_1^2, ..., \sigma_d^2)$



Example: multivariate Gaussian distribution

Multivariate Gaussian: $\boldsymbol{x} = (x_1, ..., x_d)^\top$

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp[-\frac{1}{2} \underbrace{(\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}_{:=\Delta^2}]$$

- Eigen decomposition of Σ:

$$\Sigma = U\Lambda U^\top \quad \Rightarrow \quad \Sigma^{-1} = U\Lambda^{-1}U^\top, \quad \Lambda = diag(\lambda_1, ..., \lambda_d)$$

- Define $\mathbf{y} = U^\top(\boldsymbol{x} - \boldsymbol{\mu})$

$$\Delta^2 = \mathbf{y}^\top \Lambda^{-1} \mathbf{y} = \sum_{i=1}^{d} \frac{y_i^2}{\lambda_i}$$

- Contour $\Delta^2 = C$ has an "ellipse" shape

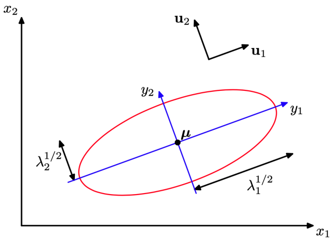


Fig from Bishop's PRML book

Going beyond Gaussian: change-of-variables rule

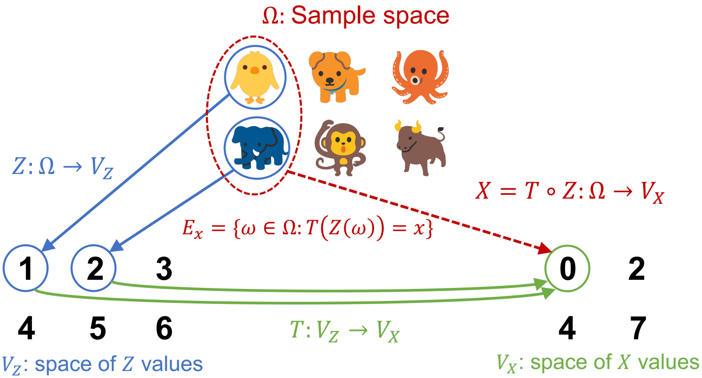
A common way to construct multivariate distribution beyond e.g., multinomial and Gaussian:

- start from random variable $Z = (Z_1, ..., Z_K)$ with distribution p_Z
- use a transformation to get $X = T(Z)$
- this induces a distribution p_X depending on T and p_Z

Q: What is the PMF/PDF of X given T and PMF/PDF of Z?

Going beyond Gaussian: change-of-variables rule

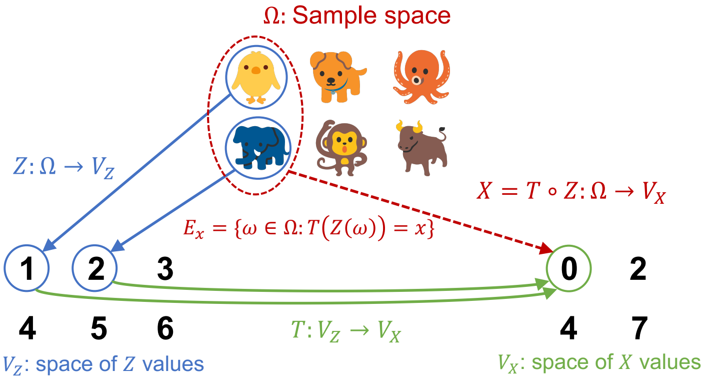
Key idea: p_X preserves the **event probability** given by $\mathbb{P}$



$$P_X(X \in \{0\}) = P_Z(Z \in \{1, 2\}) = \mathbb{P}(\omega \in \{chicken, elephant\})$$

Going beyond Gaussian: change-of-variables rule

Key idea: p_X preserves the **event probability** given by $\mathbb{P}$



$$P_X(X \in \{0\}) = P_Z(Z \in \{1, 2\}) = \mathbb{P}(\omega \in \{chicken, elephant\})$$

Discrete case: if T is invertible, then the PMF is

$$p_X(X = x) = p_Z(Z = T^{-1}(x)).$$

Going beyond Gaussian: change-of-variables rule

Key idea: p_X preserves the **event probability** given by $\mathbb{P}$

Continuous case:

- for any $x \in V_X = \mathbb{R}^{dim(X)}$, can work out $E_x = \{\omega \in \Omega : T(Z(\omega)) = x\}$
- this means for any $S \subset V_X$, can work out $E_S = \cup_{x \in S} E_x$
- note that $\mathbb{P}$ does not change!

Assume $Z : \Omega \to V_Z = \mathbb{R}^{dim(Z)}$ maps E_S to $U \subset V_Z$, then:

$$U = \{z \in V_z : T(z) \in S\} := T^{-1}(S)$$

$$\Rightarrow \quad P_X(X \in S) = P_Z(Z \in T^{-1}(S)) = \mathbb{P}(E_S)$$

$$\Rightarrow \quad \int_{\alpha \in S} p_X(X = \alpha) d\alpha = \int_{\beta \in T^{-1}(S)} p_Z(Z = \beta) d\beta$$

Going beyond Gaussian: change-of-variables rule

Key idea: p_X preserves the **event probability** given by $\mathbb{P}$

Continuous case: PDFs satisfy

$$\int_{\alpha \in S} p_X(X = \alpha) d\alpha = \int_{\beta \in T^{-1}(S)} p_Z(Z = \beta) d\beta, \; T^{-1}(S) = \{z \in V_z : T(z) \in S\}$$

For invertible and continuous T, to compute PDF p_X:

- let dz be a very small neighbourhood around z, such that $p_Z(Z = z') \approx p_Z(Z = z), \forall z' \in dz$

$$\Rightarrow \quad \int_{\beta \in dz} p_Z(Z = \beta) d\beta \approx p_Z(Z = z) dz$$

- $T(z) = x, \Rightarrow dx = T(dz)$ is also a very small neighbourhood around x, such that $p_X(X = x') \approx p_X(X = x), \forall x' \in dx$

$$\Rightarrow \quad \int_{\alpha \in dx} p_X(X = \alpha) d\alpha \approx p_X(X = x) dx$$

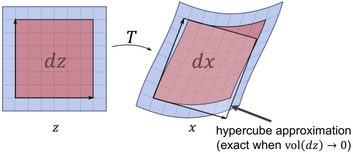
Going beyond Gaussian: change-of-variables rule

Key idea: p_X preserves the **event probability** given by $\mathbb{P}$
Continuous case: PDFs satisfy

$$\int_{\alpha \in S} p_X(X = \alpha) d\alpha = \int_{\beta \in T^{-1}(S)} p_Z(Z = \beta) d\beta, \; T^{-1}(S) = \{z \in V_z : T(z) \in S\}$$

For invertible and continuous T, to compute PDF p_X:

- Matching probability mass for the same event:
  $p_X(X = x)dx = p_Z(Z = z)dz$



$$\Rightarrow \quad p_X(X = x) = p_Z(Z = z)\left|\frac{dz}{dx}\right| = p_Z(Z = T^{-1}(x))\left|\frac{dT^{-1}(x)}{dx}\right|$$

Going beyond Gaussian: change-of-variables rule

Summary of computing PMF/PDF of X for invertible T:

- Discrete case: $p_X(X = x) = p_Z(Z = T^{-1}(x))$
- Continuous case: $p_X(X = x) = p_Z(Z = T^{-1}(x))|\frac{dT^{-1}(x)}{dx}|$

Key idea: p_X preserves the **event probability** given by $\mathbb{P}$

- Probability $\mathbb{P}$ is defined on subsets of Ω
- For $E \subset \Omega$, $U = Z(E)$ $S = T(U) = T \circ Z(E)$ are two different sets of "quantitative descriptions" of the elements in E
- So the underlying probability shouldn't change, i.e.,

$$P_X(X \in S) = P_Z(Z \in U) = \mathbb{P}(E)$$

- PMF/PDF can be work out by ensuring this match for any $E \subset \Omega$

Law of the unconscious statistician (LOTUS)

Computing expectation of X given that $X = T(Z)$:
LOTUS rule:

$$\mathbb{E}_X[f(X)] = \mathbb{E}_Z[f(T(Z))]$$

Proof for discrete case:

$$\mathbb{E}_X[f(X)] = \sum_x p_X(X = x) f(x)$$

Recall from change-of-variables rule for discrete distribution:

$$\begin{aligned} p_X(X = x) &= P_X(X \in \{x\}) = P_Z(Z \in T^{-1}(x)) \\ &= \sum_{z \in T^{-1}(x)} P_Z(Z \in z) = \sum_{z : T(z) = x} p_Z(Z = z) \end{aligned}$$

Law of the unconscious statistician (LOTUS)

Computing expectation of X given that $X = T(Z)$:
LOTUS rule:

$$\mathbb{E}_X[f(X)] = \mathbb{E}_Z[f(T(Z))]$$

Proof for discrete case:

$$\mathbb{E}_X[f(X)] = \sum_x p_X(X = x) f(x)$$

$$\Rightarrow \quad \mathbb{E}_X[f(X)] = \sum_x \left(\sum_{z:T(z)=x} p_Z(Z = z) \right) f(x)$$

$$= \sum_z p_Z(Z = z) f(T(z)) = \mathbb{E}_Z[f(T(Z))]$$

Law of the unconscious statistician (LOTUS)

Computing expectation of X given that $X = T(Z)$:
LOTUS rule:

$$\mathbb{E}_X[f(X)] = \mathbb{E}_Z[f(T(Z))]$$

Proof for continuous case, assuming T is invertible and continuous:

$$\mathbb{E}_X[f(X)] = \int p_X(x) f(x) dx$$

Recall from change-of-variables rule for continuous distribution:

$$p_X(X = x) = p_Z(Z = T^{-1}(x)) \left| \frac{dT^{-1}(x)}{dx} \right|$$

Also note that $\left|\frac{dT^{-1}(x)}{dx}\right| = \left|\frac{dz}{dx}\right|$ for $z = T^{-1}(x)$

Law of the unconscious statistician (LOTUS)

Computing expectation of X given that $X = T(Z)$:
LOTUS rule:

$$\mathbb{E}_X[f(X)] = \mathbb{E}_Z[f(T(Z))]$$

Proof for continuous case, assuming T is invertible and continuous:

$$\mathbb{E}_X[f(X)] = \int p_X(x)f(x)dx$$

$$\Rightarrow \quad \mathbb{E}_X[f(X)] = \int \left(p_Z(Z=z)|\frac{dz}{dx}|f(T(z))\right)_{z=T^{-1}(x)} dx$$

$$= \int p_Z(Z=z)f(T(z))dz = \mathbb{E}_Z[f(T(Z))]$$

Law of the unconscious statistician (LOTUS)

Computing expectation of X given that $X = T(Z)$:
LOTUS rule:

$$\mathbb{E}_X[f(X)] = \mathbb{E}_Z[f(T(Z))]$$

LOTUS is true even when T is not an invertible mapping
(The proof uses measure theory, not discussed in this course)

Summary

Today we covered:

- Multivariate probability: definition and common examples
- Change-of-variables rule
- LOTUS

Next lecture: more on multivariate probability

- vector mean and variance
- conditional distribution
- sum rule and product rule