Notations

MATH-412 - Statistical Machine Learning

Notations

x, y	plain lower case letter denote scalars or elements from general spaces
$\mathbf{x}$	bold lower case denote vectors (fixed or random depending on context)
X	capitals denote a scalar or vectorial random variable
$\{X = \mathbf{x}\}$	denotes the event that the random variable X takes the value $\mathbf{x}$
$\boldsymbol{X}$	bold capitals denote matrices (possibly random)

Expectations

Let X be a real valued random variable, and P_X the probability measure associated with X, so that for any set A.

Expectations

Let X be a real valued random variable, and P_X the probability measure associated with X, so that for any set A. Then

$$P_X(A) = \mathbb{P}(X \in A).$$

Expectations

Let X be a real valued random variable, and P_X the probability measure associated with X, so that for any set A. Then

$$P_X(A) = \mathbb{P}(X \in A).$$

The expectation is then

$$\mathbb{E}[f(X)] = \int f(x)\, dP_X(x).$$

Expectations

Let X be a real valued random variable, and P_X the probability measure associated with X, so that for any set A. Then

$$P_X(A) = \mathbb{P}(X \in A).$$

The expectation is then

$$\mathbb{E}[f(X)] = \int f(x)\, dP_X(x).$$

If X has a density p_X then

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

Expectations

Let X be a real valued random variable, and P_X the probability measure associated with X, so that for any set A. Then

$$P_X(A) = \mathbb{P}(X \in A).$$

The expectation is then

$$\mathbb{E}[f(X)] = \int f(x)\, dP_X(x).$$

If X has a density p_X then

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

Note that

$$\mathbb{P}(X \in A) = \mathbb{E}[1_{\{X \in A\}}] = \int 1_{\{x \in A\}}\, dP_X(x).$$

where $1_{\{z \in A\}}$ is the *indicator function* equal to 1 if "$z \in A$" is true and 0 else.

Conditional Expectation

Let (X, Z) a pair of random variables with joint probability distribution $P_{X,Z}$.

Conditional Expectation

Let (X, Z) a pair of random variables with joint probability distribution $P_{X,Z}$.

If (X, Z) admits a joint probability density $p_{X,Z}(x, z)$, it is possible to define

- the **marginal probability density** of Z $p_Z(z) = \int p_{X,Z}(x, z)dx$

Conditional Expectation

Let (X, Z) a pair of random variables with joint probability distribution $P_{X,Z}$.

If (X, Z) admits a joint probability density $p_{X,Z}(x, z)$, it is possible to define

- the **marginal probability density** of Z $p_Z(z) = \int p_{X,Z}(x, z)dx$
- the **conditional probability density** of X given Z as

$$p_{X|Z}(x|z) = \frac{p_{X,Z}(x, z)}{p_Z(z)},$$

Conditional Expectation

Let (X, Z) a pair of random variables with joint probability distribution $P_{X,Z}$.

If (X, Z) admits a joint probability density $p_{X,Z}(x, z)$, it is possible to define

- the **marginal probability density** of Z $p_Z(z) = \int p_{X,Z}(x, z)dx$
- the **conditional probability density** of X given Z as

$$p_{X|Z}(x|z) = \frac{p_{X,Z}(x, z)}{p_Z(z)},$$

  which is only defined for z such that $p_Z(z) > 0$.

Conditional Expectation

Let (X, Z) a pair of random variables with joint probability distribution $P_{X,Z}$.

If (X, Z) admits a joint probability density $p_{X,Z}(x, z)$, it is possible to define

- the **marginal probability density** of Z $p_Z(z) = \int p_{X,Z}(x, z)dx$
- the **conditional probability density** of X given Z as

$$p_{X|Z}(x|z) = \frac{p_{X,Z}(x, z)}{p_Z(z)},$$

  which is only defined for z such that $p_Z(z) > 0$.

In that case we can define the conditional expectation of $f(X)$ given Z as the function $z \mapsto h(z)$ defined by

$$h(z) = \mathbb{E}[f(X)|Z = z] = \int f(x)\, p_{X|Z}(x|z)\, dx.$$

Conditional Expectation

Let (X, Z) a pair of random variables with joint probability distribution $P_{X,Z}$.

If (X, Z) admits a joint probability density $p_{X,Z}(x, z)$, it is possible to define

- the **marginal probability density** of Z $p_Z(z) = \int p_{X,Z}(x, z)dx$
- the **conditional probability density** of X given Z as

$$p_{X|Z}(x|z) = \frac{p_{X,Z}(x, z)}{p_Z(z)},$$

  which is only defined for z such that $p_Z(z) > 0$.

In that case we can define the conditional expectation of $f(X)$ given Z as the function $z \mapsto h(z)$ defined by

$$h(z) = \mathbb{E}[f(X)|Z = z] = \int f(x)\, p_{X|Z}(x|z)\, dx.$$

It is often useful to consider the random variable $h(Z)$ which is also written

$$h(Z) = \mathbb{E}[f(X)|Z].$$

More general conditional distributions

In fact even when (X, Z) does not admit a joint density, under some technical assumption (e.g. if X and Z belong a finite dimensional vector space) it is possible to define a conditional probability distribution of X given Z which is denoted $P_{X|Z}$.

More general conditional distributions

In fact even when (X, Z) does not admit a joint density, under some technical assumption (e.g. if X and Z belong a finite dimensional vector space) it is possible to define a conditional probability distribution of X given Z which is denoted $P_{X|Z}$.
We then have

$$\mathbb{P}(X \in A|Z = z) = P_{X|Z}(A|z)$$

and

$$\mathbb{E}[f(X)|Z = z] = \int f(x)\, dP_{X|Z}(x|z).$$

Variance and conditional variance

Variance :

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Variance and conditional variance

Variance :

$$\mathsf{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$$

Conditional variance :

$$\mathsf{Var}(X|Z = z) = \mathbb{E}\big[(X - \mathbb{E}[X|Z = z])^2 \mid Z = z\big]$$

Variance and conditional variance

Variance :

$$\mathsf{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$$

Conditional variance :

$$\mathsf{Var}(X|Z=z) = \mathbb{E}\big[(X - \mathbb{E}[X|Z=z])^2 \mid Z=z\big]$$

$$\mathsf{Var}(X|Z) = \mathbb{E}\big[(X - \mathbb{E}[X|Z])^2 \mid Z\big]$$

Variance and conditional variance

Variance :

$$\mathsf{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$$

Conditional variance :

$$\mathsf{Var}(X|Z=z) = \mathbb{E}\big[(X - \mathbb{E}[X|Z=z])^2 \mid Z=z\big]$$

$$\mathsf{Var}(X|Z) = \mathbb{E}\big[(X - \mathbb{E}[X|Z])^2 \mid Z\big]$$

Covariance of real valued r.v.s X and Y :

$$\mathsf{cov}(X,Y) = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big]$$

Variance and conditional variance

Variance :

$$\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$$

Conditional variance :

$$\mathrm{Var}(X|Z = z) = \mathbb{E}\big[(X - \mathbb{E}[X|Z = z])^2 \mid Z = z\big]$$

$$\mathrm{Var}(X|Z) = \mathbb{E}\big[(X - \mathbb{E}[X|Z])^2 \mid Z\big]$$

Covariance of real valued r.v.s X and Y :

$$\mathrm{cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big]$$

Covariance matrix $\mathbf{C}$: If $X = (X_1, \ldots, X_p)^\top$ takes values in $\mathbb{R}^p$, we define

$$\mathbf{C} = \mathrm{Cov}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top\big],$$

with $\mathbf{C}_{ij} = \mathrm{cov}(X_i, X_j)$.

A property of conditional expectations

We have

$$\mathbb{E}[\, g(Z)\, f(X) \mid Z] = g(Z)\, \mathbb{E}[\, f(X) \mid Z]$$

A property of conditional expectations

We have

$$\mathbb{E}[\,g(Z)\,f(X) \mid Z] = g(Z)\,\mathbb{E}[\,f(X) \mid Z]$$

In words, in a conditional expectation given Z,

- all functions of Z "behave like constants",
- and so functions of Z can be "factored out".

A property of conditional expectations

We have

$$\mathbb{E}[\, g(Z)\, f(X) \mid Z] = g(Z)\, \mathbb{E}[\, f(X) \mid Z]$$

In words, in a conditional expectation given Z,

- all functions of Z “behave like constants”,
- and so functions of Z can be “factored out”.

Proof. $\mathbb{E}[\, g(Z)\, f(X) \mid Z] = h(Z)$ with

$$h(z) = \mathbb{E}[\, g(Z)\, f(X) \mid Z = z] = \mathbb{E}[\, g(z)\, f(X) \mid Z = z] = g(z)\, \mathbb{E}[\, f(X) \mid Z = z].$$