L01: the Expectation Operator

Gonzalo Ferrer

31 January 2022

1 Probability: A refresh on some useful definitions

The random variables x and y are called independent if:

$$p(x,y) = p(x)p(y).$$

More intuitively, two random variables are independent if there is no observable relation between them, for instance, the current weather in (first r.v.) and the grade you will get in the course (second r.v.).

Note: In this course $p(x)$ indicates probability density function (PDF). For a review on random variables and PDFs, check out the probability review notes.

Product rule

$$p(x,y) = p(x|y)p(y)$$

Total probability

$$p(x) = \int p(x,y)dy \qquad \text{(marginalization)}$$

$$p(x|z) = \int p(x,y|z)dy$$

Bayes Theorem

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(x,y)dx}$$

Def: Expectation Operator

$$\mathbb{E}\{x\} = \int_{-\infty}^{+\infty} x \cdot p(x)dx = \mu_x$$

2 Properties

$\mathbb{E}\{\cdot\}$ is linear

- $\mathbb{E}\{A\} = A$
- $\mathbb{E}\{Ax\} = A\,\mathbb{E}\{x\}$
- $\mathbb{E}\{A + x\} = A + \mathbb{E}\{x\}$
- $\mathbb{E}\{x + y\} = \mathbb{E}\{x\} + \mathbb{E}\{y\}$:

$$\begin{aligned}
\mathbb{E}\{x+y\} &= \int\int (x+y)p(x,y)dxdy = \\
&= \int\int xp(x,y)dxdy + \int\int yp(x,y)dxdy = \\
&= \int x\left(\int p(x,y)dy\right)dx + \int y\left(\int p(x,y)dx\right)dy = \\
&= \int xp(x)dx + \int yp(y)dy = \mathbb{E}\{x\} + \mathbb{E}\{y\}
\end{aligned}$$

3 Anti-properties

- $\mathbb{E}\{x, y\} \neq \mathbb{E}\{x\}\mathbb{E}\{y\}$ (in general)
- If x, y uncorrelated $(\sigma_{xy} = 0) \Rightarrow \mathbb{E}\{x, y\} = \mathbb{E}\{x\}\mathbb{E}\{y\}$
- If x, y independent $\underset{\not\Leftarrow}{\Rightarrow}$ x, y uncorrelated.

4 Expectation of multi-dimensional r.v.

$\mathbb{E}\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \right\} = \begin{bmatrix} \mathbb{E}\{x\} \\ \mathbb{E}\{y\} \end{bmatrix}$

5 Conditional expectation

$$\mathbb{E}\{x|y\} = \int_{-\infty}^{+\infty} x \cdot p(x|y)dx$$

6 Covariance. Scalar form

Autocovariance or covariance

$$\sigma_{xx}^2 = cov(x, x) = \mathbb{E}\big\{(x - \mathbb{E}\{x\})^2\big\} = \mathbb{E}\{x^2\} - \big(\mathbb{E}\{x\}\big)^2$$

Cross-covariance

$$\sigma_{xy}^2 = cov(x, y) = \mathbb{E}\big\{(x - \mathbb{E}\{x\})(y - \mathbb{E}\{y\})\big\}$$

7 Covariance. Vectorial form

Covariance

$$\Sigma_x = cov(x, x) = cov(x) = \mathbb{E}\big\{(x - \mathbb{E}\{x\})(x - \mathbb{E}\{x\})^\top\big\}$$

Cross-covariance

$$\Sigma_{xy} = cov(x, y) = \mathbb{E}\big\{(x - \mathbb{E}\{x\})(y - \mathbb{E}\{y\})^\top\big\}$$

Example: Expand Σ_{xy}.

$$\begin{aligned}
\Sigma_{xy} &= \mathbb{E}\big\{(x - \mathbb{E}\{x\})(y - \mathbb{E}\{y\})^\top\big\} \\
&= \mathbb{E}\big\{xy^\top + x\big(-\mathbb{E}\{y\}\big)^\top - \mathbb{E}\{x\}y^\top + \mathbb{E}\{x\}\mathbb{E}\{y\}^\top\big\} = \\
&= \mathbb{E}\{xy^\top\} - \mathbb{E}\{x\mathbb{E}\{y\}^\top\} - \mathbb{E}\big\{\mathbb{E}\{x\}y^\top\big\} + \mathbb{E}\big\{\mathbb{E}\{x\}\mathbb{E}\{y\}^\top\big\} = \\
&= \mathbb{E}\{xy^\top\} - \mathbb{E}\{x\}\mathbb{E}\{y\}^\top - \mathbb{E}\{x\}\mathbb{E}\{y^\top\} + \mathbb{E}\{x\}\mathbb{E}\{y^\top\} = \\
&= \mathbb{E}\{xy^\top\} - \mathbb{E}\{x\}\mathbb{E}\{y^\top\} = \Sigma_{xy}
\end{aligned}$$

Col 1: $\Sigma_x = \Sigma_x^\top$ (*is symmetric*. But Σ_{xy} *is* *not* *symmetric*)

$$\Sigma_x = \mathbb{E}\{xx^\top\} = \mathbb{E}\left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \right\} = \mathbb{E}\left\{ \begin{bmatrix} x_1x_1 & x_1x_2 & x_1x_3 \\ x_2x_1 & x_2x_2 & x_2x_3 \\ x_3x_1 & x_3x_2 & x_3x_3 \end{bmatrix} \right\}$$

Col 2: Σ_x is Positive (Semi)definite or PSD: $v^\top \Sigma_x v \geq 0 \; \forall v$
Proof:

$$v^\top \Sigma_x v = v^\top \mathbb{E}\big\{(x - \mathbb{E}\{x\})(x - \mathbb{E}\{x\})^\top\big\} v = \mathbb{E}\big\{ \underbrace{v^\top (x - \mathbb{E}\{x\})}_{u \text{ is now a scalar.}} (x - \mathbb{E}\{x\})^\top v \big\}$$
$$= \mathbb{E}\big\{u \cdot u\big\} \geq 0$$

8 Sample mean and sample covariance

$x_i \sim p(x)$ iid
Sample mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Sample covariance (unbiased estimate)

$$\overline{\Sigma}_x = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^\top$$