Probability theory review

Simon Kafader, 15-114-945

October 9, 2017

1.a)

If X and Y are independent, then we know that $E(XY) = E(X)E(Y)$. Thus, $Cov(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$.

Now let X be uniformly distributed in $[-1,\ 1]$ and let $Y = X^2$. Clearly, X and Y are dependent, but

$$COV[X,Y] = COV[X,X^2] = E[X \cdot X^2] - E[X]E[X^2] = E[X^3] - E[X]E[X^2] = 0 - 0 \cdot E[X^2] = 0$$

So even though the covariance here is zero, the variables are not independent.

b)

From the definition of the covariance matrix we can easily see that it must be a square matrix: $V(X) = (Cov(X_i, X_j))_{i,j=1,\ldots,n}$

Proof that it is symmetric:

$$Var[X]^\top = E[(X - E[X])(X - E[X])^\top]^\top$$

$$= E[((X - E[X])(X - E[X])^\top)^\top]$$

$$= E[(X - E[X])(X - E[X])^\top]$$

$$= Var[X]$$

Proof of semi-definite:

For vectors $x_i = (x_{i1}, ..., x_{ik})^\top, i = 1, ..., n$, the covariance matrix is $Q = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^\top$.

Let $y \in \mathbb{R}^k$be a nonzero vector:

$$y^\top Q y = y^\top \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^\top \right) y$$

$$= \frac{1}{n}\sum_{i=1}^{n} y^\top (x_i - \bar{x})(x_i - \bar{x})^\top y$$

$$= \frac{1}{n}\sum_{i=1}^{n}((x_i - \bar{x})^\top y)^\top ((x_i - \bar{x})^\top y)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left((x_i - \bar{x})^\top y\right)^2 \geq 0\,.$$

Therefore Q is always positive semi-definite.

2.)

$$E(Z) = E(AX + Y)$$

We know that the expectation is a linear function. This means we can take the matrix out and split it

$$E(AX + Y) = AE(X) + E(Y) = A \cdot 0 + \mu = \mu$$

$$Cov[Z] = Cov(Z, Z) = Cov(AX + Y, AX + Y)$$

Now since we know that the covariance is bilinear and symmetric we can split it up as follows

$$Cov(AX + Y, AX + Y) = Cov(AX, AX) + 2 \cdot Cov(AX, Y) + Cov(Y, Y)$$

$$= A^2 Cov(X, X) + 2ACov(X, Y) + Cov(Y, Y)$$

Since X and Y are independent random variables, we know that $Cov(X, Y) = 0$. Therefore

$$Cov[Z] = A^2 I + 2A0 + \sigma I = A^2 + \sigma I$$

3.) The probability of Thomas finding Viktor in this last bar is exactly $(2/3) * (1/5) = 2/15$. This is because informally Viktor decides beforehand if and if so to which party he goes. This makes the stated probability. Now because this is decided beforehand, getting new information does not change

anything about the probability since it is already fixed. Therefore it is 2/15.

4.)

We start with taking a constant out of the integral:

$$\int_{x \in \mathbb{R}^\ltimes} e^{-\frac{1}{2}x^\top Ax - x^\top b - c} = e^{-c} \int_{x \in \mathbb{R}^\ltimes} e^{-\frac{1}{2}x^\top Ax - x^\top b}$$

Now let X be an n-dimensional random variable and A its covariance matrix. Now from the hint in the excercise we know that it is normally (gaussian) distributed. Since A is positive definite the multivariate normal distribution is

$$f_{\mathbf{X}}(x_1,\ldots,x_n) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}}$$

When we use it for our remaining integral with $\mu = x - b$ we get the following equatation

$$\int_{x\in\mathbb{R}^{\mathbb{K}}} e^{-\frac{1}{2}x^\top Ax - x^\top b - c} = e^{-c}\int_{x\in\mathbb{R}^{\mathbb{K}}} e^{-\frac{1}{2}x^\top Ax - x^\top b} = \frac{\exp\left(-c - \frac{1}{2}b^\top A^{-1}b\right)}{\sqrt{(2\pi)^n|\boldsymbol{A}|}} = \frac{(2\pi)^{\frac{n}{2}}|A|^{-\frac{1}{2}}}{e^{c-\frac{1}{2}b^\top A^{-1}b}|}$$

5.) As we've seen in the lecture we prefer to use the log likelihood for maximising such functions. Therefore this log likelihood here will be:

$$l(\lambda) = \sum_{i=1}^{n}(X_i log\lambda - \lambda - log(X_i!)) = log(\lambda)\sum_{i=1}^{n}(X_i - n\lambda) - \sum_{i=1}^{n} log(X_i!)$$

$$l'(\lambda) = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0$$

This implies that the estimate should be $\lambda = \overline{X}$

6.)

Let's start at the right side:

$$f(x_1)\cdot f(x_2|\,x_1)\cdot\cdots\cdot f(x_n|\,x_1,\ldots x_{n-1}) = f(x_1)\cdot\frac{f(x_1,x_2)}{f(x_1)}\cdot\frac{f(x_1,x_2,x_3)}{f(x_1,x_2)}\cdot\cdots\cdot\frac{f(x_1,\ldots,x_n)}{f(x_1,\ldots,x_{n-1})}$$

Now we can strike away the numerator i with the denominator i+1 what only leaves one term:

$$f(x_1)\cdot\frac{f(x_1,x_2)}{f(x_1)}\cdot\frac{f(x_1,x_2,x_3)}{f(x_1,x_2)}\cdot\cdots\cdot\frac{f(x_1,\ldots,x_n)}{f(x_1,\ldots,x_{n-1})} = f(x_1,\ldots,x_n)$$