TRINITY COLLGE DUBLIN

School of Computer Science and Statistics

Extra Questions ST3009: Statistical Methods for Computer Science

Question 1. Suppose two continuous valued random variables X and Y have the following joint PDF

$$f_{XY}(x,y) = \begin{cases} 0 & x < 0, y < 0 \\ 1 & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & x > 1, y > 1 \end{cases}$$

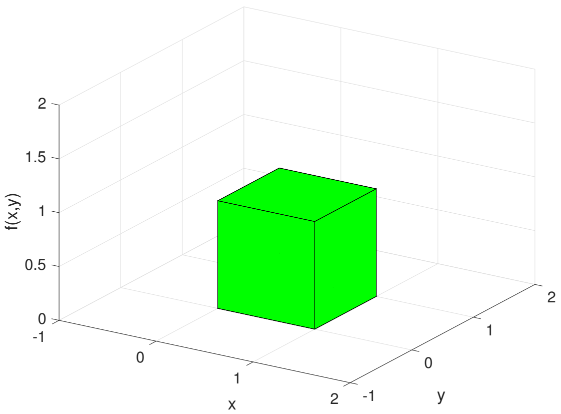

Figure 1: Plot of PDF $f_{XY}(x,y)$

(a) Calculate $P(0 \le X \le 0.5 \text{ and } 0 \le Y \le 0.5)$?
(b) Calculate $P(0 \le X \le 2 \text{ and } 0 \le Y \le 0.5)$?
Solution

- $P(0 \le X \le 0.5 \text{ and } 0 \le Y \le 0.5)$ is the volume under the PDF for $0 \le X \le 0.5$ and $0 \le Y \le 0.5$. This is a cube of length 0.5, breadth 0.5 and height 1, so its volume is $0.5 \times 0.5 \times 1 = 0.25$.

- We split the volume under the PDF for $0 \le X \le 2$ and $0 \le Y \le 0.5$ into two parts, the volume under under the PDF for $0 \le X \le 1$ and $0 \le Y \le 0.5$ and the volume $1 \le X \le 2$ and $0 \le Y \le 0.5$. The volume under the PDF for $0 \le X \le 1$ and $0 \le Y \le 0.5$ is a cube of length 1, breadth 0.5 and height 1 so its volume is $1 \times 0.5 \times 1 = 0.5$. The volume under the PDF for $1 \le X \le 1$ and $0 \le Y \le 0.5$ is 0. So the total volume is 0.5 i.e. $P(0 \le X \le 2 \text{ and } 0 \le Y \le 0.5) = 0.5$

Question 2. Suppose two continuous valued random variables X and Y have the following joint CDF

$$F(x,y) = \begin{cases} 0 & x < 0, y < 0 \\ xy & 0 \le x \le 1, 0 \le y \le 1 \\ y, x > 1, 0 \le y \le 1 & \\ x, y > 1, 0 \le x \le 1 & \\ 1 & x > 1, y > 1 \end{cases}$$

(a) Sketch the graph of this CDF.
(b) Calculate $P(X \leq 0.5 \text{ and } Y \leq 0.5)$
(c) Calculate $P(0.1 \leq X \leq 0.5 \text{ and } 0.1 \leq Y \leq 0.5)$?
(d) Calculate $P(0 \leq X \leq 2 \text{ and } 0 \leq Y \leq 0.5)$?
(e) Are X and Y independent ? Hint: recall $P(X \leq x) = F(x, \infty)$.
Solution

- $P(X \leq 0.5 \text{ and } Y \leq 0.5) = F(0.5, 0.5) = 0.25$
- $P(0 \leq X \leq 0.5 \text{ and } 0 \leq Y \leq 0.5) = F(0.5, 0.5) - F(0.1, 0.1) = 0.25 - 0.01 = 0.024$
- $P(0 \leq X \leq 2 \text{ and } 0 \leq Y \leq 0.5) = F(2, 0.5) - F(0, 0) = 0.5 - 0 = 0.5$
- For independence we need $P(X \leq x \text{ and } Y \leq y) = P(X \leq x)P(Y \leq y)$. Now $P(X \leq x) = F(x, \infty) = x$ for $0 \leq x \leq 1$ and $P(Y \leq y) = F(\infty, y) = y$ for $0 \leq y \leq 1$, so $P(X \leq x)P(Y \leq y) = xy$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$. But $P(X \leq x \text{ and } Y \leq y) = F(x, y) = xy$ when $0 \leq x \leq 1$ and $0 \leq y \leq 1$. That is, $P(X \leq x \text{ and } Y \leq y) = P(X \leq x)P(Y \leq y)$. We can repeat for $x < 0$ and $x > 1$ etc to confirm that this holds for all values of x and y. Hence, X and Y are independent.

Question 3. Suppose two continuous valued random variables X and Y have the following joint CDF

$$F(x,y) = \begin{cases} 0 & x < 0, y < 0 \\ x^2y/2 + xy^3/2 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ x^2/2 + x/2 & 0 \leq 0 \leq x \leq 1, y \geq 1 \\ y/2 + y^3/2, x > 1, 0 \leq y \leq 1 & \\ 1 & x > 1, y > 1 \end{cases}$$

(a) Calculate $P(X \leq 0.5 \text{ and } Y \leq 0.5)$
(b) Calculate $P(0.1 \leq X \leq 0.5 \text{ and } 0.1 \leq Y \leq 0.5)$?
(c) Calculate $P(0 \leq X \leq 2 \text{ and } 0 \leq Y \leq 0.5)$?
Solution

- $P(X \leq 0.5 \text{ and } Y \leq 0.5) = F(0.5, 0.5) = 0.5^2 \times 0.5/2 + 0.5 \times 0.5^3/2 = 0.0938$
- $P(0 \leq X \leq 0.5 \text{ and } 0 \leq Y \leq 0.5) = F(0.5, 0.5) - F(0.1, 0.1) = 0.0938 - 0.00055 = 0.0932$
- $P(0 \leq X \leq 2 \text{ and } 0 \leq Y \leq 0.5) = F(2, 0.5) - F(0, 0) = 0.5/2 + 0.5^3/2 - 0 = 0.3125$

Question 4. Suppose random variables X and Y have PDFs $f_X(x) = e^{-x}$, $f_Y(y) = 0.5e^{-0.5y}$. Suppose also the X and Y are independent.
(a) What is their joint PDF ?
(b) Sketch a graph of this PDF.
Solution

- Since they are independent we just multiply the individual PDFs. That is, $f_{XY}(x, y) = 0.5e^{-x}e^{-0.5y} = 0.5e^{-x-0.5y}$

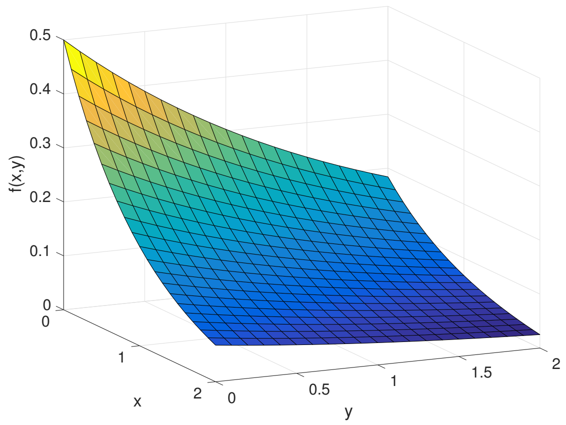

Figure 2: Plot of PDF $f_{XY}(x, y)$

Question 5. Suppose random variable X had PDF

$$f_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 & 0 < x \leq 1 \\ 0 & x > 1 \end{cases}$$

and random variable Y has PDF

$$f_Y(y) = \begin{cases} 0 & y \leq 0.5 \\ 1 & 0.5 < y \leq 1.5 \\ 0 & y > 1.5 \end{cases}$$

Suppose also the X and Y are independent.

(a) What is their joint PDF ?

(b) Sketch a graph of this PDF.

Solution

- Since they are independent we just multiply the individual PDFs. That is,

$$f_{XY}(x, y) = \begin{cases} 0 & x \leq 0 \text{ or } y \leq 0.5 \\ 1 & 0 < x \leq 1, 0.5 < y \leq 1.5 \\ 0 & x > 1 \text{ or } y > 1.5 \end{cases}$$

Question 6. Suppose two random variables X and Y have PDFs $f_X(x) = e^{-x}$, $f_Y(y) = 0.5e^{-0.5y}$ and conditional PDF $f_{Y|X}(y|x) = e^{-|x-y|}$. Using Bayes Rule for PDFs write an expression for $f_{X|Y}(x|y)$.

Solution

- $f_{X|Y}(x|y) = f_{Y|X}(y|x)f_X(x)/f_Y(y) = e^{-|x-y|}e^{-x}/(0.5e^{-y}) = 2e^{-|x-y|-x+y}$

Question 7. (a) Give Bayes rule for PDFs

(b) Explain the difference between the maximum likelihood and the MAP estimate of a random variable

(c) Suppose after observing data the likelihood of parameter θ is $L(\theta) = e^{-(\theta-1)^2}$. What is the maximum likelihood estimate of θ ?

Solution

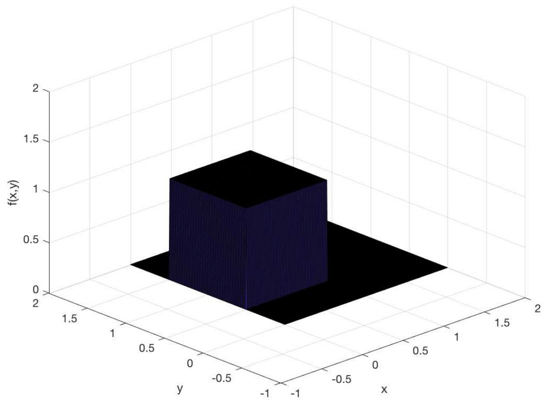

Figure 3: Plot of PDF $f_{XY}(x, y)$

- The value of θ which maximises $e^{-(\theta-1)^2}$ is $\theta = 1$

Question 8. Suppose and urn contains balls and that fraction θ of the balls are white and the rest are red. I draw n balls, with replacement, from the urn and let X be the number of white balls observed.

(a) Give an expression for the likelihood $P(X = x|\theta)$

(b) Suppose $n = 100$ and I observe 25 white balls. What is the maximum likelihood estimate for θ (use matlab to plot the value of $P(X = x|\theta)$ for a range of values of θ).

(c) Suppose now that before drawing the balls my prior probability was $P(\theta) = \frac{1}{20\pi}e^{-100(\theta-0.5)^2}$ and for simplicity assume that $P(X = 25) = 1$ (since it just scales the posterior). Give an expression for the posterior $P(\theta|X = x)$ (use Bayes rule).

(d) What is the MAP estimate for θ (use matlab to plot the value of $P(\theta|X = x)$ for a range of values of θ). Discuss why it differs from the maximum likelihood estimate.

Solution

- The probability of drawing x white balls is $P(X = x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$.
- The maximum likelihood estimate is $\theta = 0.25$
- The posterior is $P(\theta|X = x) = P(X = x|\theta)P(\theta)/P(X = x) = \frac{1}{20\pi}\binom{n}{x}\theta^x(1-\theta)^{n-x}e^{-100(\theta-0.5)^2}$.
- The MAP estimate is approximate $\theta = 0.32$. The prior says that we believe $\theta = 0.5$ with high probability. After observing the data we change our belief to a lower value, but because of the prior its still higher than the maximum likelihood. As the number n of balls drawn is increased the two estimates will, however, converge to the same value.