MAST90105 Lab and Workshop 6 Solutions

The Lab and Workshop this week cover problems arising from Module 3, Section 4 (Normal Distribution and Central Limit Theorem), Module 4 and Module 5, Section 1.

1 Lab

The objective of this week is to use Mathematica to do actual calculations on moment generating functions and probability density functions for sums of independent random variables.

1. Suppose $X_1 \stackrel{d}{=} N(\mu = 3, \sigma^2 = 4)$, $X_2 \stackrel{d}{=} N(3, 4)$, and X_1 and X_2 are independent.

   a. Use Mathematica to find the mgf of X_1 and X_2.
      - *See Lab6Answers.nb, 1(a)*

   b. Let $Y = 5X_1 - 2X_2 + 6$. Find the mgf of Y using Mathematica.
      - *See Lab6Answers.nb, 1(b)*

   c. Name the distribution of Y and give the values of the associated parameters.
      - $5X_1 - 2X_2 + 6 \sim N(15, 58)$. *(Note: Mathematica uses standard deviation as the second parameter of the normal distribution, but the textbook and notes use variance, as the answer given here has.)*

2. Let $\bar{X}$ be the mean of a random sample of size 12 from the uniform distribution on the interval (0,1) (which has the mean 1/2 and variance 1/12). Use Mathematica to do the following:

   a. Find the pdf of $\bar{X}$. The command to get the pdf of the sum of 12 independent uniform random variables on (0,1) is `PDF[UniformSumDistribution[12], x]` and plot it - see Lab6.nb for some help on this - why does the help work?
      - *Let* $X = \sum_{i=1}^{12} U_i$ *where* $U_1, \cdots, U_{12}$ *are independent uniform random variables on the unit interval. Let* $F_{\bar{X}}, F_X$ ($f_{\bar{X}}, f_X$) *be the cdf's (respectively, pdf's) of* $\bar{X}, X$. *Then* $F_{barX}(y) = P(X/12 \le y) = P(X \le 12y) = F_X(12y)$. *Differentiating gives* $f_{\bar{X}}(y) = 12 f_X(12y)$,. *See Lab6Answers.nb, 2(a).*
      - *Note that* $/.\{x \to 12y\}$ *has the effect of replacing the variable* x *with* $12y$ *and the plot is over* y *between 0 and 1.*

   b. Find the probability $P(1/2 \le \bar{X} \le 2/3)$ based on the exact distribution of $\bar{X}$. Translate the required probability into one about the sum of the 12 independent uniform random variables.
      - $P(1/2 \le \bar{X} \le 2/3) = P(1/2 \le X/12 \le 2/3) = P(6 \le X \le 8)$
        $= F_X(8) - F_X(6)$. *See Lab6Answers.nb, 2(b).*

   c. By the CLT $\bar{X}$ approximately has a normal distribution with mean 1/2 and variance 1/144. Now approximate the probability $P(1/2 \le \bar{X} \le 2/3)$ based on this normal distribution. Compare the result with that in (b).

- *See Lab6Answers.nb, 2(c).*

3. Let $Y = A + B + C + D$ be the sum of a random sample of size 4 from the distribution whose pdf is that of $U^{\frac{1}{3}}$ where $U \sim U(0, 1)$.

   a. The function $t1$ in the notebook Lab6.nb sets up the pdf of $U^{\frac{1}{3}}$. Plot it and find the mean and variance of A.

   - *See Lab6Answers.nb, 3(a).*

   b. Find the pdf of Y, and plot the pdf of Y.

   - *See Lab6Answers.nb, 3(b).*

   c. Calculate $P(0.3 \leq Y \leq 1.5)$ based on the pdf of Y.

   - *See Lab6Answers.nb, 3(c)*

   d. Approximate $P(0.3 \leq Y \leq 1.5)$ based on the normal distribution, and see how close it is to the result of (c).

   - *See Lab6Answers.nb, 3(d)*
   - *The approximation is not so good because the interval* $[0.3, 1.5]$ *is well away from the centre of the distribution around 3, so it is the very small values of both the exact and approximate densities that are being used in both probability calculations.*

4. Alter the code from Lab 4 Question 3 to investigate in Mathematica the relationship between the Gamma distribution with large values of α and the normal distribution. Take the scale parameter to be 1. How will you adapt the distance D?

   - *See See Lab6Answers.nb, 4.*
   - *If* f, g *are densities, the same argument as used in Lab 4 shows that* $\max_A |P(A) - Q(A)| = \int_{-\infty}^{\infty} |f(x) - g(x)|\, dx$ *if* P, Q *are the probabilities from the densities* f, g.
   - *Note that numerical integration is needed because symbolic integration is too slow and accuracy restriction is needed as well.*
   - *It helps to note that there is little probability outside the mean plus or minus five standard deviations for both distributions so that the range of integration and plotting can be restricted.*

2 Workshop

You can use Mathematica or R for numerical calculations as needed.

5. Suppose $X_1 \stackrel{d}{=} N(\mu = 3, \sigma^2 = 4)$, $X_2 \stackrel{d}{=} N(3, 4)$, and X_1 and X_2 are independent.

   a. Write down the mgf of X_1 and X_2.

   - $M_{X_1}(t) = M_{X_2}(t) = e^{3t+2t^2}$

b. Let $Y = 5X_1 - 2X_2 + 6$. Find the mgf of Y.

- $E_Y(t) = E(e^{t(5X_1 - 2X_2 + 6)}) = e^{6t}E(e^{5tX})E(e^{-2tX_2})$
  $= e^{6t}M_{X_1}(5t)M_{X_2}(-2t) = e^{6t}e^{15t+50t^2}e^{-6t+8t^2} = e^{15t+58t^2}$

c. Name the distribution of Y and give the values of the associated parameters.

- *Therefore* $Y \stackrel{d}{=} N(15, 116)$.

6. The serum zinc level X in micrograms per deciliter for males between ages 15 and 17 has a distribution which is approximately normal with $\mu = 90$ and $\sigma = 15$. Compute the conditional probability $P(X > 120|X > 105)$.

- $P(X > 120|X > 105) = \frac{P(X>120)}{P(X>105)} = \frac{1-\Phi(2)}{1-\Phi(1)} = \frac{0.0288}{0.1587} = 0.1437$.

7. Let $Z_1, Z_2, \cdots, Z_7$ be a random sample from the standard normal distribution $N(0, 1)$. Let $W = Z_1^2 + Z_2^2 + \cdots + Z_7^2$. Name the distribution of W with associated parameter value. Justify your answer.

- $W \stackrel{d}{=} \chi^2(7)$. *This is because a squared standard normal random variable has a* $\chi^2(1)$ *distribution, and the sum of independent chi-square random variables is chi-square with the degrees of freedom being the sum of the individual degrees of freedom of the random variables.*

8. If $X_1, X_2, \cdots, X_{16}$ is a random sample of size $n = 16$ from the normal distribution $N(50, 100)$.

a. What is the distribution of $\frac{1}{100}\sum_{i=1}^{16}(X_i - 50)^2$?

- $\chi^2(16)$

b. What is the distribution of $\bar{X}$?

- $N(50, \frac{100}{16} = 6.25)$.

9. Let $\bar{X}$ be the mean of a random sample of size 12 from the uniform distribution on the interval (0,1) (which has the mean 1/2 and variance 1/12). Approximate the probability $P(1/2 \le \bar{X} \le 2/3)$ using the central limit theorem.

- *By CLT,* $\bar{X} \stackrel{d}{\approx} N(\frac{1}{2}, \frac{1}{144})$.
- *So* $P(\frac{1}{2} \le \bar{X} \le \frac{2}{3}) \approx P(\frac{1/2-1/2}{1/12} \le Z \le \frac{2/3-1/2}{1/12})$
  $= P(0 \le Z \le 2) = \Phi(2) - \Phi(0) = 0.9772 - 0.4 = 0.4772$.

10. Let the distribution of Y be $Binomial(25, 1/2)$. Find the probability $P(10 \le Y \le 12)$ in two ways: using the binomial pmf formula, and using the normal approximation. Comment on any difference.

- $P(10 \le Y \le 12) = \sum_{y=10}^{12}\binom{25}{y}0.5^{25} = 0.3852$.

- *By CLT,* $Y \overset{d}{\approx} N(12.5, 6.25)$.
  So $P(10 \le Y \le 12) = P(9.5 \le Y \le 12.5) \approx P(\frac{9.5-12.5}{2.5} \le Z \le \frac{12.5-12.5}{2.5})$
  $= P(-1.2 \le Z \le 0) = \Phi(0) - \Phi(-1.2) = 0.5 - 0.1151 = 0.3849$.

11. The number X of flaws on a certain tape of length one yard follows a Poisson distribution with mean 0.3. We examine $n = 100$ such tapes and count the total number Y of flaws.

  a. Assuming independence, what is the distribution of Y?

  - $Y \overset{d}{=} \text{Poisson}(30)$. *By CLT* $Y \overset{d}{\approx} N(30, 30)$.

  b. Find the exact and approximate probabilities for $P(Y \le 25)$.

  - $P(Y \le 25) = P(Y \le 25.5) \approx P(Z \le \frac{25.5-30}{\sqrt{30}}) = P(Z \le -0.8216) = \Phi(-0.8216) = 0.2057$.

12. Let $Y = X_1 + X_2 + \cdots + X_{15}$ be the sum of a random sample of size 15 from the distribution whose pdf is $f(x) = (3/2)x^2,\ -1 < x < 1$. Approximate $P(-0.3 \le Y \le 1.5)$ using the central limit theorem.

  - $E(X_1) = \int_{-1}^{1} (3/2)x^3 dx = 0$. $\text{Var}(X_1) = \int_{-1}^{1} (x-0)^2 (3/2)x^2 dx = \frac{3}{5}$.
  - *By CLT* $Y \overset{d}{\approx} N(15 \cdot 0, 15 \cdot \frac{3}{5}) = N(0, 9)$.
  - *So* $P(-0.3 \le Y \le 1.5) \approx P(\frac{-0.3-0}{3} \le Z \le \frac{1.5-0}{3}) = P(-0.1 \le Z \le 0.5) = 0.2313$.

13. Let the joint pmf of X and Y be defined by

$$f(x,y) = \frac{x+y}{32}, \quad x = 1,2,\ y = 1,2,3,4.$$

  a. Find the marginal pmf of X.

  - $f_X(x) = \sum_{y=1}^{4} \frac{x+y}{32} = \frac{x+1}{32} + \frac{x+2}{32} + \frac{x+3}{32} + \frac{x+4}{32} = \frac{4x+10}{32}, \quad x = 1, 2.$

  b. Find the marginal pmf of Y.

  - $f_Y(y) = \sum_{x=1}^{2} \frac{x+y}{32} = \frac{1+y}{32} + \frac{2+y}{32} = \frac{3+2y}{32}, \quad y = 1,2,3,4.$

  c. Calculate $P(X > Y)$.

  - $P(X > Y) = P(\{X = 2, Y = 1\}) = \frac{2+1}{32} = \frac{3}{32}$.

  d. Calculate $P(Y = 2X)$.

  - $P(Y = 2X) = P(\{X = 1, Y = 2\} \cup \{X = 2, Y = 4\}) = \frac{1+2}{32} + \frac{2+4}{32} = \frac{9}{32}$.

  e. Calculate $P(X + Y = 3)$.

  - $P(X + Y = 3) = P(\{X = 1, Y = 2\} \cup \{X = 2, Y = 1\}) = \frac{1+2}{32} + \frac{2+1}{32} = \frac{6}{32}$.

  f. Calculate $P(X \le 3 - Y)$.

  - $P(X \le 3 - Y) = P(X + Y \le 3) = P(\{X = 1, Y = 1\} \cup \{X = 1, Y = 2\} \cup \{X = 2, Y = 1\}) = \frac{1+1}{32} + \frac{1+2}{32} + \frac{2+1}{32} = \frac{8}{32} = \frac{1}{4}$.

g. Are X and Y independent?

- *No, because* $f(x,y) \neq f_X(x)f_Y(y)$.

h. Find $E(X)$.

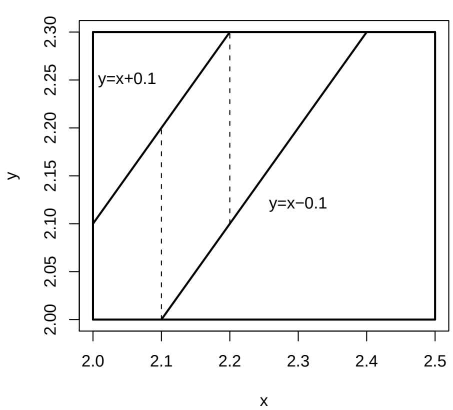
- $E(X) = \sum_{x=1}^{2} x\frac{4x+10}{32} = \frac{4+10}{32} + \frac{36}{32} = \frac{50}{32} = \frac{25}{16}$.

i. Find $E(X+Y)$.

- $E(X+Y) = \sum_{x=1}^{2}\sum_{y=1}^{4}(x+y)\frac{x+y}{32}$
$= \frac{(1+1)^2+(1+2)^2+(1+3)^2+(1+4)^2+(2+1)^2+(2+2)^2+(2+3)^2+(2+4)^2}{32} = \frac{140}{32}$.

14. Two construction companies make bids of X and Y (in \$100,000's) on a remodeling project. The joint pmf of X and Y is uniform on the space $2 < x < 2.5$, $2 < y < 2.3$. If X and Y are within 0.1 of each other, the companies will be asked to rebid; otherwise the lower bidder will be awarded the contract. What is the probability that they will be asked to rebid?

- $P(|X-Y| < 0.1) = P(X - 0.1 < Y < X + 0.1)$
$= \int_{2.0}^{2.1}\int_{2.0}^{x+0.1} \frac{1}{(2.5-2)(2.3-2)}dydx + \int_{2.1}^{2.2}\int_{x-0.1}^{x+0.1} \frac{1}{(2.5-2)(2.3-2)}dydx$
$+ \int_{2.2}^{2.4}\int_{x-0.1}^{2.3} \frac{1}{(2.5-2)(2.3-2)}dydx$
$= \frac{(x-1.9)^2}{0.3}\Big|_{2.0}^{2.1} + \frac{0.2\times 0.1}{0.15} - \frac{(x-2.4)^2}{0.3}\Big|_{2.2}^{2.4} = \frac{11}{30}$.
- *In above calculation, note that* $2 < X < 2.5$ *and* $2 < Y < 2.3$ *are implicitly required.*
- *Because the joint pdf is uniform, this probability can also be obtained by calculating the proportion of the relevant area on the support of* X *and* Y*.*



15. Let $f(x,y) = 2e^{-x-y}$, $0 \leq x \leq y < \infty$, be the joint pdf of X and Y.

a. Find the marginal pdf $f_X(x)$ of X.

- $f_X(x) = \int_x^{\infty} 2e^{-x-y}dy = 2e^{-2x}, \quad 0 \leq x < \infty$.

b. Find the marginal pdf $f_Y(y)$ of Y.

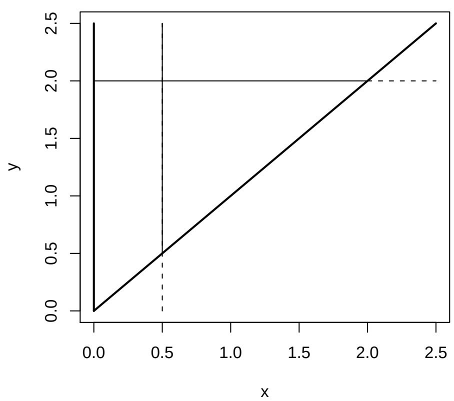
- $f_Y(y) = \int_0^y 2e^{-x-y}dx = 2e^{-y}(1-e^{-y}), \quad 0 \le y < \infty.$

c. Compute $E(X)$ and $E(e^{-X-2Y})$.

- $E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^{\infty} x \cdot 2e^{-2x}dx = \frac{1}{2}$. *Or*
  $E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)dydx = \int_0^{\infty}\int_x^{\infty} x \cdot 2e^{-x-y}dydx = \frac{1}{2}.$
- $E(e^{-X-2Y}) = \int_0^{\infty}\int_x^{\infty} e^{-x-2y} \cdot 2e^{-x-y}dydx$
  $= \int_0^{\infty}\int_x^{\infty} 2e^{-2x-3y}dydx = \int_0^{\infty} \frac{2}{3}e^{5x}dx = \frac{2}{15}.$

d. Compute $P(X > \frac{1}{2})$.

- $P(X > \frac{1}{2}) = \int_{1/2}^{\infty} f_X(x)dx = \int_{1/2}^{\infty} 2e^{-2x}dx = e^{-1}.$



e. Compute $P(X > \frac{1}{2}, Y > 2)$.

- $P(X > \frac{1}{2}, Y > 2) = \int_2^{\infty}\int_{1/2}^{\infty} f(x,y)dxdy = \int_2^{\infty}\int_{1/2}^{y} 2e^{-x-y}dxdy$
  $= \int_2^{\infty}(2e^{-\frac{1}{2}-y} - 2e^{-2y})dy = [-2e^{-\frac{1}{2}-y} + e^{-2y}]_2^{\infty} = 2e^{-2.5} - e^{-4}.$

f. Compute $P(Y > 2|X > \frac{1}{2})$.

- $P(Y > 2|X > \frac{1}{2}) = \frac{P(Y>2, X>\frac{1}{2})}{P(X>\frac{1}{2})} = \frac{2e^{-2.5}-e^{-4}}{e^{-1}} = 2e^{-1.5} - e^{-3}.$

16. Let the joint pmf of X and Y be

$$f(x,y) = \frac{1}{4}, \quad (x,y) \in S = \{(0,0),(1,1),(1,-1),(2.0)\}.$$

a. Represent the joint pmf by a table.

X \ Y	-1	0	1	$f_X(x)$
0		$\frac{1}{4}$		$\frac{1}{4}$
1	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{2}$
2		$\frac{1}{4}$		$\frac{1}{4}$
$f_Y(y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

b. Are X and Y independent?

- *No, because the space of X and Y is not rectangular.*

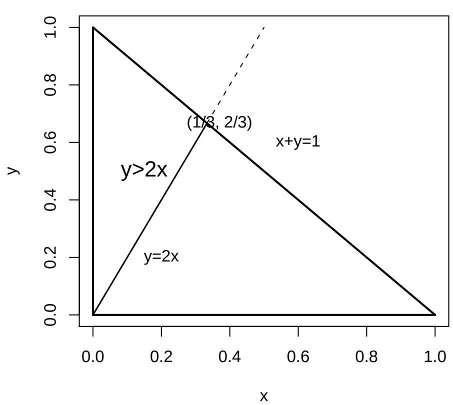
c. Calculate $\mathrm{Cov}(X, Y)$ and ρ.

- *From the marginal pmf's in the table in (a),* $\mu_X = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1.$
  $\mu_Y = (-1) \times \frac{1}{4} + 0 \times \frac{1}{2} + 1 \times \frac{1}{4} = 0.$
  $\sigma_X^2 = (0-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{1}{2} + (2-1)^2 \times \frac{1}{4} = \frac{1}{2}.$
  $\sigma_Y^2 = (-1)^2 \times \frac{1}{4} + 0^2 \times \frac{1}{2} + 1^2 \times \frac{1}{4} = \frac{1}{2}.$
- $E(XY) = 0 \times 0 \times \frac{1}{4} + 1 \times (-1) \times \frac{1}{4} + 1 \times 1 \times \frac{1}{4} + 2 \times 0 \times \frac{1}{4} = 0.$
- *So* $\mathrm{Cov}(X, Y) = E(XY) - \mu_X \mu_Y = 0 - 1 \times 0 = 0.$
  And $\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = 0.$
- *This is another example of dependent variables having zero correlation coefficient.*

17. Consider continuous random variables X and Y which have the following joint pdf

$$f(x, y) = 24xy, \quad x > 0,\ y > 0,\ x + y < 1.$$

a. Sketch a graph of the support of X and Y.



b. Find the probability $P(Y > 2X)$.

- $P(Y > 2X) = \int_0^{1/3} \int_{2x}^{1-x} 24xydydx = \frac{7}{27}$.

c. Find the marginal pdf $f_1(x)$ of X.

- $f_1(x) = \int_0^{1-x} 24xydy = 12x(1-x)^2, \quad 0 < x < 1$.

d. Find the mean $E(X)$.

- $E(X) = \int_0^1 x12x(1-x)^2 dx = \frac{2}{5}$.

e. Find the variance $\mathrm{Var}(X)$.

- $\mathrm{Var}(X) = \int_0^1 (x - \frac{2}{5})^2 12x(1-x)^2 dx = \frac{1}{25}$.

f. Find the covariance $\mathrm{Cov}(X, Y)$.

- $\mathrm{Cov}(X, Y) = \int_0^1 \int_0^{1-y} (x - \frac{2}{5})(y - \frac{2}{5}) 24xydxdy = -\frac{2}{75}$

g. Find the correlation coefficient ρ between X and Y.

- $\rho = \frac{-2/75}{\sqrt{1/25}\sqrt{1/25}} = -\frac{2}{3}$

h. Find the conditional pdf $h(y|x)$ of Y given $X = x$.

- $h(y|x) = \frac{24xy}{12x(1-x)^2} = \frac{2y}{(1-x)^2}, \quad 0 < y < 1 - x$.

i. Find the condition probability $P(Y \le \frac{1}{3}(1-X)|X = x)$.

- $P(Y \le \frac{1}{3}(1-X)|X = x) = \int_0^{(1-x)/3} h(y|x)dy = \frac{1}{9}$.

j. Find the conditional mean $E(Y|X = x)$.

- $E(Y|x) = \int_{-\infty}^{\infty} y \cdot h(y|x)dy = \int_0^{1-x} y \frac{2y}{(1-x)^2} dy = \frac{2(1-x)}{3}$.

18. Show that $\mathrm{Cov}(aX + b, cX + d) = ac\mathrm{Var}(X)$, where X is a random variable and a, b, c are deterministic constants.

- $\mathrm{Cov}(aX + b, cX + d) = E[(aX + b - E[aX + b])(cX + d - E[cX + d])]$
  $= E[(aX - E[aX])(cX - E[cX])] = E[ac(X - E[X])(X - E[X])]$
  $= acE[(X - E[X])^2] = ac\mathrm{Var}(X)$.

19. Let the pmf of X be $f_1(x) = \frac{1}{10}$, $x = 0, 1, 2, \cdots, 9$, and the conditional pmf of Y given $X = x$ be $h(y|x) = \frac{1}{10-x}$, $y = x, x+1, \cdots, 9$. Find

a. the joint pmf $f(x, y)$ of X and Y.

- $f(x, y) = f_1(x)h(y|x) = \frac{1}{10(10-x)}, \quad x = 0, 1, 2, \cdots, 9, y = x, x+1, \cdots, 9$.

b. The marginal pmf $f_2(y)$ of Y.

- $f_2(y) = \sum_{x=0}^{y} \frac{1}{10(10-x)}, \quad y = 0, 1, 2, \cdots, 9$.

c. $E(Y|x)$.

- $E(Y|x) = \sum_{y=x}^{9} y \cdot h(y|x) = \sum_{y=x}^{9} \frac{y}{10-x} = \frac{(x+9)(9-x+1)}{2(10-x)} = \frac{x+9}{2}$.

20. The marginal distribution of X is $U(0,1)$. The conditional distribution of Y, given $X=x$, is $U(0,e^x)$.

    a. Determine $h(y|x)$, the conditional pdf of Y, given $X=x$.

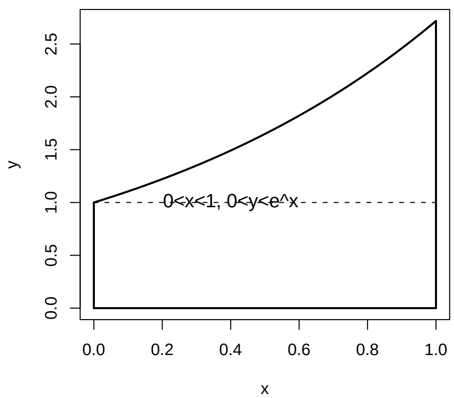
    - $h(y|x) = \frac{1}{e^x} = e^{-x}, \quad 0 < y < e^x.$

    b. Find $E(Y|x)$.

    - $E(Y|x) = \frac{e^x}{2}, \quad 0 < x < 1.$

    c. Find the joint pdf of X and Y. Sketch the region where $f(x,y) > 0$.

    - $f(x,y) = f_1(x)h(y|x) = 1 \times e^{-x} = e^{-x}, \quad 0 < x < 1, 0 < y < e^x.$



    d. Find $f_2(y)$, the marginal pdf of Y.

    - $f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx = \begin{cases} \int_0^1 e^{-x}dx = 1 - e^{-1}, & \text{if } 0 < y \le 1 \\ \int_{\ln(y)}^1 e^{-x}dx = y^{-1} - e^{-1}, & \text{if } 1 < y < e \end{cases}$

    e. Find $g(x|y)$, the conditional pdf of X, given $Y=y$.

    - $g(x|y) = \frac{f(x,y)}{f_2(y)} = \begin{cases} \frac{e^{-x}}{1-e^{-1}}, & 0 < x < 1, & \text{if } 0 < y \le 1 \\ \frac{e^{-x}}{y^{-1}-e^{-1}}, & \ln(y) < x < 1, & \text{if } 1 < y < e \end{cases}.$

21. An obstetrician does ultrasound examinations on her patients between their 16th and 25th weeks of pregnancy to check on the growth of the unborn child. Let X equal the widest diameter of the head and Y be the length of the femur, both in mm. Assume that X and Y have a bivariate normal distribution with $\mu_X = 60.6$, $\sigma_X = 11.2$, $\mu_Y = 46.8$, $\sigma_Y = 8.4$, $\rho = 0.94$

    a. Find $P(40.5 < Y < 48.9)$

    - *The marginal distribution of* Y *is* $N(46.8, 8.4^2)$ *so* $P(40.5 < Y < 48.9) = P(Y \le 48.9) - P(Y \le 40.5) = 0.3721(4dp)$ *can be found from the R commands*

```
pnorm(48.9,mean=46.8, sd = 8.4)-pnorm(40.5,mean=46.8, sd = 8.4)

## [1] 0.372079
```

b. Find $P(40.5 < Y < 48.9|X = 68.6)$

- *The conditional distribution of* $Y|X = 68.6$ *is* $N(\mu_Y+\rho\sigma_Y z_X(68.6), (1-\rho^2)\sigma_Y^2)$ *so the required probability is 0. 1084 (4dp) from the R commands*

```
  (meanygivenx = 46.8 + 0.94*8.4*(68.6-60.6)/11.2)

## [1] 52.44

  (sigmaygivenx = (1-0.94^2)^0.5*8.4)

## [1] 2.865865

  pnorm(40.5,mean=meanygivenx, sd = sigmaygivenx)

## [1] 1.548046e-05
```

22. Karl Pearson carried out a famous study on the resemblances between fathers and sons. He found that the distribution from his sample of 1078 pairs of fathers and sons had a mean and standard deviation of height for fathers of 69 and 2 (in.), whereas for sons it was 70 and 2 (in.). The correlation between fathers' and sons' heights was 0.5. Assuming that the distribution can of the pairs of heights can be modelled as bivariate normal with the sample means, sds and correlation, what is

a. the expected height for the son of a father who is 74 in. tall,

- *Let* (X, Y) *be the heights (in.) of the father and son, so that* (X, Y) *has a bivariate normal distribution with the parameters* $\mu_X = 69, \sigma_X = 2, \mu_Y = 70, \sigma_T = 2, \rho = 0.5$.
- *Using the argument in Module 4,* $E(Y|X = 74) = \mu_Y + \rho\sigma_Y z_X(74) = 70 + 0.5 \times 2 \times \frac{74-69}{2} = 72.5$ *so the expected height is 72.5 in.*

b. the chance that the son's height is more than 1 in. from the expected height in (a)?

- *The conditional distribution of* $Y|X = 74$ *is* $N(\mu_Y + \rho\sigma_Y z_X(74), (1 - \rho^2)\sigma_Y^2)$ *so the required probability is (4dp) from the R commands*

```
(meanygivenx = 70 + 0.5*2*(74-69)/2)

## [1] 72.5

(sigmaygivenx = (1-0.5^2)^0.5*2)

## [1] 1.732051
```

```
# Note that more than 1 inch different from 72.5 is < 71.5 or > 73.5.
# Symmetry of the normal pdf about the mean
# implies the two probs have equal probability
2*pnorm(71.5,mean=meanygivenx, sd = sigmaygivenx)

## [1] 0.5637029
```

c. the expected height of a father whose son is 72.5 in?

- *Using the argument in Module 4,* $E(X|Y = 72.5) = \mu_X + \rho\sigma_X z_Y(72.5) = 69 + 0.5 \times 2 \times \frac{72.5-70}{2} = 70.25$ *so the expected height is 70.25 in.*

d. the chance that both a father and son are above average height?

- *This one requires a subtle argument.*
- *From lectures the standardised* (X, Y) *satisfes:*

$$(z_X(X), z_Y(Y)) = (Z_1, \rho Z_1 + \sqrt{1-\rho^2} Z_2),$$

where (Z_1, Z_2) *have independent* $N(0, 1)$ *distributions.*

- *So the required probability is the same probability as*

$$P(z_X(X) \geq 0, z_Y(Y) \geq 0) = P(Z_1 \geq 0, \rho Z_1 + \sqrt{1-\rho^2} Z_2 \geq 0).$$

- *This last probability is the integral of the joint density of* (Z_1, Z_2) *over the set* $A = \{(z_1, z_2) : z_1 \geq 0, z_2 \geq \frac{-\rho z_1}{\sqrt{1-\rho^2}}\}$.
- *The set* A *is* $\frac{1}{3}$ *of the plane in the sense that if it is rotated through* $120°$ *three times it comes back to itself - draw a diagram and use some trigonometry to see this. Specifically,*

$$\mathbb{R}^2 = A \cup A_1 \cup A_2$$

where A_1 *is the rotation of* A *by* $120°$ *and* A_2 *is a rotation of* A *by* $240\circ$.

- *To see this note that* A *consists of the positive quadrant plus a wedge of the quadrant with non-negative* z_1 *and* z_2 *bounded by the line* $z_2 = -\frac{\rho}{\sqrt{1-\rho^2}} z_1 = -\frac{1}{\sqrt{3}} z_1$. *The angle that this line makes with the* z_1-*axis is* $30°$ *(since it's* $\tan = -\frac{1}{\sqrt{3}}$*) , so the angle of the boundary of* A *at the origin is* $90+30 = 120°$.
- *Since the joint density of* (Z_1, Z_2) *is not affected by rotations, the probability is* $\frac{1}{3}$.
- *To see this in detail, if* ϕ *is the joint density of* (Z_1, Z_2) *then*

$$\begin{aligned} 1 &= \int_A \phi(x, y)\, dx\, dy + \int_{A_1} \phi(x, y)\, dx\, dy + \int_{A_2} \phi(x, y)\, dx\, dy \\ &= 3 \int_A \phi(x, y)\, dxdy, \end{aligned} \tag{1}$$

and $3 \int_A \phi(x, y)\, dxdy$ *is the required probability.*

- *The second equality in 1 follows because the rotational symmetry means that* $\phi(x, y) = \phi(x_1, y_1) = \phi(x_2, y_2)$ *where* (x_1, y_1) *is the rotation of* (x, y) *by* $120°$ *and* (x_2, y_2) *is the rotation of* (x, y) *by* $240°$ *so*

$$\int_A \phi(x, y)\, dx\, dy = \int_{A_1} \phi(x, y)\, dx\, dy = \int_{A_2} \phi(x, y)\, dx\, dy.$$

23. Scores in the two exams in a double weight course have means 65 and 60, standard deviations 18 and 20 and a correlation of 0.75. Assuming they can be modelled as bivariate normal, what is the expected score in the second exam for a student how is above average on the first exam?

    - *72 is the answer rounded to the nearest integer.*
    - *To see this the result of question 24 is used.*
    - *To see this, let* Z_1, Z_2 *be the random variables giving the scores in the first and second exams in statistical units.*
    - *From question 24 with* $b = \infty, a = 0$, $E(Z_2|Z_1 > 0) = \frac{0.75[1-0]}{\sqrt{2\pi}(1/2}$ *and the answer,* z*, is given in the following R code.*
    - *To get the expected second exam score,* x *for an above average student in the first exam note first that* $Z_1 > 0$ *is the same event as the first exam score being above average. So the required expected score,* x *is obtained by converting* z *from statistical units back to the original scale. This is done in the R code.*

```
( z =  0.75*2/sqrt(2*pi))

## [1] 0.5984134

(x = 60 + 20*z)

## [1] 71.96827
```

24. Suppose (X, Y) has a standard bivariate normal density with correlation ρ. For a, b, find $E(Y|a < X < b)$.

    - $E(Y|X = x) = \rho x$.
    - $f_X(x|a < X < b) = \frac{\phi(x)}{\Phi(b)-\Phi(a)}$ *for* $a < x < b$.
    - *So* $E(Y|a < X < b) = \int_a^b \rho x \frac{\phi(x)}{\Phi(b)-\Phi(a)}\, dx = \frac{\rho[e^{-a^2/2} - e^{-b^2/2}]}{\sqrt{2\pi}(\Phi(b)-\Phi(a)}$, *on changing variables in the integal from* x *to* $\frac{x^2}{2}$.

25. Suppose (X, Y) are random variables and a, b, c, d are constants. Show

$$Cov(aX + bY, cX + dY) = acVar(X) + bdVar(Y) + (bc + da)Cov(X, Y).$$

$$\begin{aligned}
Cov(aX+bY, cX+dY) &= E((aX+bY)(cX+dY) - E(aX+bY)E(cX+dY) \\
&= E((ad+bc)XY + acX^2 + bdY^2) \\
&- (aE(X)+bE(Y))(cE(X)+dE(Y)) \\
&= (ad+bc)(E(XY) - E(X)E(Y)) \\
&+ ac(E(X^2) - (E(X))^2) + bd(E(Y^2) - (E(Y))^2) \\
&= acVar(X) + bdVar(Y) + (bc+da)Cov(X,Y).
\end{aligned}$$

26. Suppose that $X \sim N(0,1)$ and that Z is an independent random variable which takes the values $1, -1$ each with probability $\frac{1}{2}$.

a. Show that $X, -X$ have the same distribution.

$$\begin{aligned}
P(X \le x) &= \int_{-\infty}^{x} \phi(z)\,dz \\
&= \int_{-\infty}^{x} \phi(-z)\,dz \\
&= \int_{-x}^{-\infty} \phi(y)\,dy \quad (y = -z) \\
&= P(X \ge -x) \\
&= P(-X \le x),
\end{aligned}$$

as required.

b. What is the distribution of ZX?

This is the standard normal because

$$\begin{aligned}
P(ZX \le x) &= P(X \le x, Z = 1) + P(-X \le x, Z = -1) \\
&= P(X \le x)P(Z = 1) + P(-X \le x)P(Z = -1) \\
&= P(X \le x)(P(Z = 1) + P(Z = -1)) \\
&= P(X\ le - x).
\end{aligned}$$

c. Find $Cov(X, ZX), Corr(ZX, X)$.

- *Since X, ZX both have variance 1, $Cov(X, ZX) = Corr(ZX, X)$.*
- *Since X, ZX both mave mean 0 and Z is independent of X,*

$$Cov(X, ZX) = E(XZX) = E(X^2)E(Z) = 0,$$

with the last step using $E(Z) = -1 \times 1/2 + 1 \times 1/2 = 0$.

d. Are X and ZX independent? (Hint: consider $P(ZX \le -1 | X \le -1)$.)

Thare not independent because the conditional probability would have to be $P(ZX \leq -1)$ which is the same as $P(X \leq -1)$. But

$$\begin{aligned} P(ZX \leq -1 | X \leq -1) &= \frac{P(X \leq -1, Z = 1, X \leq -1) + +P(-X \leq -1, Z = -1, X \leq -1)}{P(X \leq -1)} \\ &= \frac{P(X \leq -1)}{2P(X \leq -1)} \\ &= \frac{1}{2} & \neq \end{aligned}$$

e. Is (X, Y) bivariate normal?
Although the marginal distributions are both normal, (X, Y) cannot be bivariate normal because they are uncorrelated but not independent.

27. Suppose X has a uniform pdf on [0,1]. Find the pdf of $-\ln(1-U)$.

- *For $x > 0$, $P(-\ln(1-U) \leq x) = P(U \leq 1-e^{-x})$ so $-\ln(1-U)$ has an exponential distribution with rate and scale $= 1$.*
- *Note that $p \to \ln(1-p)$ is the quantile function, so this is an example of the random number generation in Section 1 of Mondule 5.*

28. Suppose X has the Laplace pdf, f, given by $f(x) = \frac{1}{2}e^{-|x|}$ for all real x.

a. Find the cdf of X(hint: consider negative x first and compare with the exponential density).

- *For $x \leq 0$, $P(X \leq x) = \int_{-\infty}^{x} \frac{1}{2}e^{z}\,dz = \frac{1}{2}e^{-|x|}$.*
- *For $x > 0$, $P(X \leq x) = \frac{1}{2} + \int_{0}^{x} \frac{1}{2}e^{-z}\,dz = 1 - \frac{1}{2}e^{-|x|}$.*
- *So the cdf, F_X, is given by*

$$F_X(x) = \begin{cases} \frac{1}{2}e^{-|x|} & x \leq 0 \\ 1 - \frac{1}{2}e^{-|x|} & x > 0 \end{cases}.$$

b. Find the cdf of X^2.

- *For $x > 0$,*

$$\begin{aligned} P(X^2 \leq x) &= P(-\sqrt{x} \leq X \leq \sqrt{x}) \\ &= F_X(\sqrt{x}) - F_X(-\sqrt{x}) \\ &= 1 - \frac{1}{2}e^{-\sqrt{x}} - \frac{1}{2}e^{-\sqrt{x}} \\ &= 1 - e^{-\sqrt{x}}. \end{aligned}$$

- *So the cdf, F_{X^2}, is given by*

$$F_{X^2}(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\sqrt{x}} & x > 0 \end{cases}.$$

c. Find the pdf of X^2.

Differentiating the pdf, f_{X^2}, is given by

$$f_{X^2}(x) = \begin{cases} 0 & x \leq 0 \\ \frac{e^{-\sqrt{x}}}{2\sqrt{x}} & x > 0 \end{cases}.$$

d. Find the cdf of X^4.

- *For $x > 0$,*

$$\begin{aligned} P(X^4 \leq x) &= P(-\sqrt[4]{x} \leq X \leq \sqrt[4]{x}) \\ &= F_X(\sqrt[4]{x}) - F_X(-\sqrt[4]{x}) \\ &= 1 - \frac{1}{2}e^{-\sqrt[4]{x}} - \frac{1}{2}e^{-\sqrt[4]{x}} \\ &= 1 - e^{-\sqrt[4]{x}}. \end{aligned}$$

- *So the cdf, F_{X^4}, is given by*

$$F_{X^4}(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\sqrt[4]{x}} & x > 0 \end{cases}.$$

e. Find the pdf of X^4.

Differentiating the pdf, f_{X^4}, is given by

$$f_{X^4}(x) = \begin{cases} 0 & x \leq 0 \\ \frac{e^{-\sqrt[4]{x}}}{2\sqrt[4]{x}} & x > 0 \end{cases}.$$