Assignment sheet 2

September 26, 2018

Assignment 1. Let $X, Y \sim \mathcal{N}(0,1)$ independent. Show that $X+Y$ is independent of $X-Y$.
(Hints : Use transformation theorem for multivariate density/distributions.
(a) Define $U = X+Y$ and $V = X-Y$ and compute their joint density $f_{U,V}(u,v)$.
(b) Recognize the distributions U and V and prove independence.)

Assignment 2. This exercise will focus again on correlation and dependence.
(1) Let $X \sim \mathcal{N}(0,1)$.
(a) Compute $\mathbb{E}(X^2)$, $\mathbb{E}(X^3)$, $\mathbb{E}(X^4)$.
(b) Compute $\mathrm{Corr}(X, X^2)$. Discuss the result.
(2) (optional)
Use the commands

```
set.seed(27092017)
x=rnorm(500)
x2=x^2
y=cbind(x,x2)
cov(y)
cor(y)
plot(x,x2)
```

Try the above code with sample sizes 1000 and 10000. Discuss the results.

Assignment 3. There are many relations between the Poisson and the binomial distribution. We give three such examples in this exercise.
(a) The yearly number of car accidents (X) in the canton of Vaud can be modelled by a Poisson distribution (see below why). In a given accident, the probability of a casualty is p. Let us find the distribution of the number of car accidents with casualties (Y).
Let $X \sim \mathit{Poisson}(\lambda)$ and let Y be $\mathit{Binom}(X,p)$ conditional upon X.
(a) Find the joint distribution of X and Y.
(b) Find the marginal distribution of Y. Does it belong to a family of distributions you know ?
(*Hint : there will be an infinite sum that you can calculate upon noticing its elements correspond to a Poisson distribution.*)
(c) Find the conditional expectation and variance of Y given X and verify that $\mathrm{Var}\, Y = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}[\mathbb{E}(Y|X)]$.
(d) Let $X' \sim \mathit{Poisson}(\mu)$ represent the number of bicycle accidents, and assume that it is independent of X. Find the distribution of the total number of accidents $X + X'$.
(*Hint : use the moment generating function.*)
(e) What is the distribution of the number of bicycle accidents if we know that the total number accidents is k ?
(f) (optional)
Run the following commands :

```
lambda <- 7
n <- 180
x <- 0:(6*lambda)
plot(x, dbinom(x, size = n, prob = lambda/n), ylab = "")
points(x, dpois(x, lambda), col = "red")
```

Interpret what you see. Try changing the value of n.
(This phenomenon is called *the law of rare events*. There are many (namely, n) car journeys, and the probability of having an accident in each journey is small (namely, λ/n). The total number of car accidents is therefore $\text{Binom}(n, \lambda/n)$, and it can be approximated by a $\text{Poisson}(\lambda)$ distribution.)
(g) Recall that $(1 + x/n)^n \to e^x$ for any $x \in \mathbb{R}$, as $n \to \infty$. Find the limit of the moment generating function M_{B_n} for $B_n \sim Binom(n, \lambda/n)$. Can you identify the limit function as a moment generating function of a known distribution?

Assignment 4. The goal of this exercise is to show that the Geometrical distribution is the only (discrete) memoryless distribution.
Let $X \sim \text{Geom}(p)$.

(a) Show that $\mathbb{P}(X \geq k) = (1 - p)^k$.

(b) Prove the lack of memory property for the Geometrical distribution, i.e. that

$$\mathbb{P}(X \geq k + m | X \geq k) = \mathbb{P}(X \geq m).$$

(c) Show that if Y is a random variable with values in $\{0\} \cup \mathbb{N}$ respecting the lack of memory property, then Y follows a Geometrical distribution.
(*Hint : Write* $\mathbb{P}(Y \geq k) = \mathbb{P}(Y \geq k + 1 | Y \geq 1)$ *to find the c.d.f. of a Geometrical distribution.*)

(Notes :
(1) Intuitively : say you are gambling at the roulette. The lack of memory of the geometric implies that it makes no sense to implement a betting strategy based on observation of what numbers came out in the past. If your lucky number is 14 and in the first n rounds it never came out, the probability that it won't come out until the $(n+m)$-th round it doesn't depend on n (that is, the number of rounds already gone) but only on the m rounds still to come.
(2) In the continuous case, the same fact holds true for the Exponential distribution, as mentioned on Slide 63.)

Assignment 5. In this exercise, we will derive the marginal and the conditional densities associated with a bivariate normal distribution. Suppose that $\mathbf{X} = (X_1, X_2)^T$ has a bivariate normal density with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$, where $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in (0, \infty)$ and $\rho \in (-1, 1)$. We will write $\mathbf{X} \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. The joint density function is given by

$$\begin{aligned} & f_{\mathbf{X}}(x_1, x_2) \\ = \; & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \, exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \right] \right\}, \end{aligned}$$

where $x_1, x_2 \in \mathbb{R}$.
(a) Find the marginal density function of X_1.
(*Hint : To integrate out* x_2*, try to adjust the expression in the exponential so as to get a square form in* x_2*, i.e., a term like* $(x_2 - ax_1 - b)^2$*. Then, relate this with the density of some normal distribution (with appropriate variance) and use the fact that the density of a normal distribution integrates to one.*)
(b) Can you figure out the marginal density function of X_2 without any calculation?
(c) Can you identify the marginal densities?
(d) Find the conditional density function of X_1 given $X_2 = x_2$ and that of X_2 given $X_1 = x_1$.

(e) Are the conditional densities normal ?
(f) Find $\mathbb{E}[X_1 \mid X_2]$ and $\mathbb{E}[X_2 \mid X_1]$.
(Note : So, in this case, the best approximation of X_1, in terms of mean square, using a function of X_2 is actually a linear function in X_2. The vice-versa is also true. Such statements are not true in general for other distributions.)
(g) Verify that $\mathbb{E}\{\mathbb{E}[X_1 \mid X_2]\} = \mathbb{E}[X_1]$.
(h) Use the properties of conditional expectations to find $\mathrm{Cov}(X_1, X_2)$.
(i) Write down the mean vector and the covariance matrix of $\mathbf{X}$.
(j) Find $\mathrm{Var}[X_1 \mid X_2]$ and $\mathrm{Var}[X_2 \mid X_1]$.
(k) Verify that $\mathrm{Var}(X_2) = \mathbb{E}[\mathrm{Var}(X_2 \mid X_1)] + \mathrm{Var}(\mathbb{E}[X_2 \mid X_1])$.

Assignment 6. (optional)
(a) Use the command

```
set.seed(27092017)
n = 1000
rpois(n,2)
```

to generate $n = 1000$ i.i.d. samples from a Poisson(2) distribution. Store the values as a vector X. Find the mean of X and store the value as mean_est.
(b) Run the following code :

```
mean_out = 0
M = 10
for (j in 0:M)
{
mean_out = mean_out + j*dpois(j,2)
}
```

What is the value of mean_out ? Does the answer change if M is set to 100 ? Why ? Interpret your result.
(c) Run the following code :

```
mean_out1 = 0
M = 10
for (k in 0:M)
{
mean_out1 = mean_out1 + (1-ppois(k,2))
}
```

What is the value of mean_out1 ? How does it compare with mean_out and mean_est ?
(d) We will now see the theoretical justification of the previous question.
Show that if Y is a random variable taking only non-negative integer values such that $\mathbb{E}[Y] < \infty$, then $\mathbb{E}[Y] = \sum_{k=0}^{\infty} \mathbb{P}[Y > k]$.
(Hint : Write $\mathbb{P}[Y > k] = \sum_{j=k+1}^{\infty} \mathbb{P}[Y = j]$ and count how many times $\mathbb{P}[Y = j_0]$ occurs in the sum $\sum_{k=0}^{\infty} \mathbb{P}[Y > k]$ for each $j_0 \in \{0, 1, 2, \ldots\}$.)
(e) Use the command

```
set.seed(27092017)
n = 1000
rgamma(n,shape=3,rate=2)
```

to generate $n = 1000$ i.i.d. samples from a Gamma(3,2) distribution. Store the values as a vector X.
(f) Run the following code :

```
mean_out = 0
for (i in 1:n)
{
mean_out = mean_out + X[i]/n
}
```

Is the code calculating the expected value of the empirical cumulative distribution of X ? What is the value of mean_out ? What is the expected value of a random variable having the Gamma(3,2) distribution ?
(g) Run the following code :

```
mean_outs = rep(0,10)
for (j in 1:10)
{
set.seed(27092017)
n = 5^j
X = rgamma(n,shape=3,rate=2)
mean_outs[j] = mean(X)
}
plot(c(1:10),mean_outs,type="o",xaxp=c(1,10,9),xlab=expression(log[5](n)),
ylab="Est. mean")
abline(a=1.5,b=0,col="red")
```

What do you observe ? What does it tell you about the behaviour of the sample mean as the sample size increases ?
(Note : This phenomenon is called the *Law of Large Numbers.*)
(h) Write a user-defined function that will compute `1 - pgamma(x,shape=3,rate=2)`. Denote this function by Sfun.
(i) Run the command :

```
mean_new = integrate(Sfun,0,Inf)
```

What is the value of mean_new ? How does it compare with the elements of the vector mean_outs and the expected value of the Gamma(3,2) distribution ?
(Note : For a random variable Y taking values in $(0, \infty)$ with finite expectation, it can be proved that $\mathbb{E}[Y] = \int_0^\infty \{1 - F(y)\} dy$, where F is the cumulative distribution function of Y.)