

ASSIGNMENT SHEET 3

Spring 2025

Assignment 1. Prove that the following distributions are members of an exponential families by finding their natural and their usual parametrisation.

- (i) The Poisson distribution.
- (ii) The Geometric distribution.
- (iii) The exponential distribution.
- (iv) The Gamma distribution.

Assignment 2. Using the factorisation theorem, find a sufficient statistics for an independent sample $Y = (Y_1, \dots, Y_n)$ from the following distributions :

- (i) The Poisson distribution.
- (ii) The Geometric distribution.
- (iii) The exponential distribution.
- (iv) The Gamma distribution with parameters (λ, r) for λ assumed to be a known number.

You can check your results using your knowledge from Exponential Families.

Assignment 3. In this assignment, we will find the distribution of the maximum and the minimum *order statistics* for a sample.

Let X_1, X_2, \dots, X_n be an i.i.d. sample from a probability distribution with c.d.f. F and density function f . Let $X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$ and $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$.

- (a) Find the c.d.f.'s of $X_{(1)}$. Hence find the density function of $X_{(1)}$.
(Hint : $\mathbb{P}[X_{(1)} > y] = \mathbb{P}[X_1 > y, X_2 > y, \dots, X_n > y]$ since the minimum is larger than y iff all sample values are larger than y . Now compute $\mathbb{P}[X_{(1)} \leq y]$ using the formula $1 - \mathbb{P}[X_{(1)} > y]$.)
- (b) Find the c.d.f.'s of $X_{(n)}$. Hence find the density function of $X_{(n)}$.
(Hint : $\mathbb{P}[X_{(n)} \leq z] = \mathbb{P}[X_1 \leq z, X_2 \leq z, \dots, X_n \leq z]$ since the maximum is smaller than or equal to z iff all sample values are smaller than or equal to z .)
- (c) Find the joint c.d.f. of $\mathbf{W} = (X_{(1)}, X_{(n)})^\top$.
(Hint : First compute $\mathbb{P}[X_{(1)} > y, X_{(n)} \leq z]$ using arguments as in (a) and (b). Then, compute $\mathbb{P}[X_{(1)} \leq y, X_{(n)} \leq z]$ using the formula $\mathbb{P}[X_{(n)} \leq z] - \mathbb{P}[X_{(1)} > y, X_{(n)} \leq z]$. Be CAREFUL about the ranges of the variables y and z .)
- (d) Use part (c) to find the joint density function of $\mathbf{W} = (X_{(1)}, X_{(n)})^\top$. Are $X_{(1)}$ and $X_{(n)}$ independent?
- (e) Compute the marginal c.d.f.s and density functions of $X_{(1)}$ and $X_{(n)}$ for the $\text{Unif}(0, \theta)$ distribution. Also find the joint density function of $X_{(1)}$ and $X_{(n)}$.
- (f) Discuss the behaviour of the marginal c.d.f. of $X_{(n)}$ in part (e) when $n \rightarrow \infty$.
(Hint : Write the c.d.f. for z ranging on the whole of \mathbb{R} and then find the limit. Be CAREFUL at the point $z = \theta$.)

Assignment 4. Let V be a $N(\mu, \sigma^2)$ variable.

- (a) Find the c.d.f. and the density function of V^2 directly by computing $\mathbb{P}[V^2 \leq w]$.
(Note : This is an example of a non-monotone transformation of variable, and we solve it directly.)
- (b) In case $\mu = 0$ and $\sigma^2 = 1$, can you identify the distribution obtained in (a)?
- (c) Let U be a $N(\theta, \tau^2)$ variable that is independent of V . Prove that U and V^2 are independent by showing that

$$\mathbb{P}[U \leq u, V^2 \leq w] = \mathbb{P}[U \leq u] \mathbb{P}[V^2 \leq w].$$

(Note : It can be shown that, more generally, if U and V are independent, then so are $f(U)$ and $g(V)$, where f and g are two “measurable” functions.)

Let X_1, X_2 be i.i.d. $N(\gamma, \eta^2)$ variables. Let \bar{X} and S^2 be the sample mean and the sample variance.

(d) Show that $S^2 = (X_1 - X_2)^2/2$.

(e) Using part (c) and Assignment 1 in Week 2, show that \bar{X} and S^2 are independent.

(f) In case $\gamma = 0$, what is the distribution of $T = (X_1 + X_2)/|X_1 - X_2|$?

(g) (optional) Run the following command :

```
vals = matrix(0,nrow=1000,ncol=2)
n = 100
for (i in 1:1000)
{
  set.seed(i+04102017)
  X = rnorm(n)
  vals[i,1] = mean(X)
  vals[i,2] = var(X)
}
plot(vals[,1],vals[,2],xlab=expression(bar(X)),ylab=expression(S^2))
abline(v=0,col="red")
abline(h=1,col="red")
```

Try to guess the value of $Cov(\bar{X}, S^2)$ from the plot. Justify your answer.

Assignment 5. In this assignment we shall see how the entropy is related to information. Let X be a *discrete* random variable.

(a) Show that $H(X) \geq 0$. *Note : $H(X)$ can be infinite, but this happens for rather pathological distributions.*

(b) Let g be an injective function. Show that $H(g(X)) = H(X)$. What this means is that the entropy does not see what *values* X takes, but only their *probabilities*.

(c) Let X take the values $-1, 0, 1$ with probabilities $0 < p_1, p_2, p_3$, $p_1 + p_2 + p_3 = 1$. Show that $H(X^2) < H(X)$. The entropy decreased because we lost information about X : we only know its absolute value but not the sign.

Hint : the function $h(x) = x \log x$ is superadditive : $h(x+y) > h(x) + h(y)$ if $x, y > 0$.

Remark : the result holds more generally : if g is not injective (and $H(X)$ is finite), then $H(g(X)) < H(X)$.

(d) The situation in the continuous case is far from being obvious. Suppose that $X \sim \text{Unif}[0, \theta]$ for $\theta > 0$. Find $H(X)$.

(e) Is it true that $H(X) \geq 0$?

(f) Is it true that $H(g(X)) \leq H(X)$? What if g is injective? *Hint : take $\theta = 1$ and g linear.*

Assignment 6. (Optional)

In this assignment we shall see how the t distribution behaves when the number of degrees of freedom changes.

(a) Use the commands

```
set.seed(04102017)
N <- 1e5
rt(N, df = 1)
```

to generate a sample of size 100'000 from a t distribution with one degree of freedom. Store the values as a vector X .

(b) Use the command

```
sum(-1 < X & 1 > X)
```

to compute how many elements of X are in $(-1,1)$. Modify the command to obtain the *proportion* of elements of X that are in $(-1,1)$.

(c) Repeat this for 2, 4, 12, 25, 100, 250, 500 degrees of freedom. Store the resulting proportions in a vector `small` of length 8. *Hint : you can do this in an automated way with the following commands.*

```
df <- c(1, 2, 4, 12, 25, 100, 250, 500)
small <- numeric(length(df))
```

Then use a for loop that will set each element in `small` using the commands in (a) and (b).

Remark. The use of `numeric(length(df))` instead of simply putting 8 makes the code more easily modifiable : if you change the length of the vector `df` then the length of `small` adapts automatically.

(d) Do the same for a random variable $N(0,1)$, and store the proportion as a variable `small.norm`.

(e) Plot the values of `small` as a function of the degrees of freedom. Compare with the value of `small.norm`. What do you observe? *Hint : after you draw the plot you can use the following command (modify it appropriately) :*

```
abline(h = 0.63)
```

(f) Use the commands

```
xval <- seq(from = -6, to = 6, length.out = 1000)
plot(xval, dnorm(xval), type = "l", xlab = "", ylab = "")
lines(xval, dt(xval, df = 1))
```

To simultaneously plot the densities of the $N(0,1)$ and t_1 distributions. Use an appropriate for loop for plotting the densities corresponding to the different degrees of freedom in (c). What do you observe? *Hint : you can add `col = i` to the command `lines` so that different degrees of freedom will have different colours. It may be helpful to redraw the normal density again after the for loop.*