M378K Introduction to Mathematical Statistics
Homework assignment #4

Please, provide your **complete solutions** to the following problems. Final answers only, even if correct will earn zero points for those problems.

Problem 4.1. *(10 points) Let X be a continuous random variable with the cumulative distribution function denoted by F_X and the probability density function denoted by f_X.*

Express the cumulative distribution function and the density of the random variable $\tilde{X} = X^2$ in terms of F_X and f_X.

Solution: The range/support of $\tilde{X}$ is within $[0, \infty)$. For every $x \geq 0$, the cumulative distribution function is

$$\begin{aligned} F_{\tilde{X}}(x) = \mathbb{P}[\tilde{X} \leq x] = \mathbb{P}[X^2 \leq x] &= \mathbb{P}[X \leq \sqrt{x}] - \mathbb{P}[X < -\sqrt{x}] \\ &= \mathbb{P}[X \leq \sqrt{x}] - \mathbb{P}[X \leq -\sqrt{x}] = F_X(\sqrt{x}) - F_X(-\sqrt{x}). \end{aligned}$$

As for the probability density function, we have that for all $x > 0$,

$$f_{\tilde{X}}(x) = F'_{\tilde{X}}(x) = \frac{1}{2\sqrt{x}}(f_X(\sqrt{x}) + f_X(-\sqrt{x})).$$

Problem 4.2. *(10 points) Let Y be lognormal with parameters $\mu = 1$ and $\sigma = 2$, i.e., let $Y \overset{(d)}{=} e^X$ with $X \sim N(\mu, \sigma)$.*

Define $\tilde{Y} = 3Y$.

Find the median of $\tilde{Y}$, i.e., find the value m such that $\mathbb{P}[\tilde{Y} \leq m] = 1/2$.

Solution: It can be shown that $\tilde{Y}$ is lognormal with parameters $\mu^* = \mu + \ln(3)$ and $\sigma^* = \sigma$. So, $\tilde{Y}$ can be written as $\tilde{Y} = e^{\tilde{X}}$ where $\tilde{X} \sim N(\mu^*, (\sigma^*)^2)$. Hence, with m denoting the median of $\tilde{Y}$, we have

$$\begin{aligned} 1/2 &= \mathbb{P}[\tilde{Y} \leq m] \\ &= \mathbb{P}[e^{\tilde{X}} \leq m] \\ &= \mathbb{P}[\tilde{X} \leq \ln(m)]. \end{aligned}$$

Since $\tilde{X}$ is normal with mean μ^* (and the mean and the median of a normal r.v. are one and the same), we conclude that

$$\ln(m) = 1 + \ln(3) \quad \Rightarrow \quad m = 3e \approx 8.15.$$

Problem 4.3. *(10 points) Let T denote the time for a call center employee to respond to any single telephone call. We model the random variable T by uniform distribution on the interval $(48, 72)$ with the time being measured in seconds. Let R denote the* **rate** *at which the call center employee responds to queries expressed in the number of customers* per minute.

Does the random variable R have a density? If so, find the density of R.

Solution:

(i)

$$f_T(t) = \frac{1}{24}\mathbb{I}_{(48,72)} = \begin{cases} \frac{1}{24} & \text{for } 48 < x < 72 \\ 0 & \text{otherwise.} \end{cases}$$

(ii)

$$R = 60/T.$$

(iii)

$$(60/72, 60/48) = (5/6, 5/4)$$

(iv) Method of transformations: For $5/6 < r < 5/4$,

$$f_R(r) = \frac{1}{24} \times \frac{60}{r^2} = \frac{5}{2r^2}\,.$$

Problem 4.4. *(20 points) Let X, Y and Z be independent and uniformly distributed on $(0, 1)$. Find the density function of $W = X + Y + Z$.*

Solution:

Method I: The cdf-method. Let us introduce the random variable $T = X + Y$. We already showed in class that its density function looks like this

$$f_T(a) = \begin{cases} a & \text{for } a \in [0,1] \\ 2 - a & \text{for } a \in [1,2] \\ 0 & \text{otherwise.} \end{cases}$$

Next, we note that $W = T + Z$. Let's look at its cdf.

$$\begin{aligned} F_W(w) = \mathbb{P}[W \le w] = \mathbb{P}[T + Z \le w] &= \int\int f_{T,Z}(t,z)\mathbf{1}_{\{0 \le t+z \le w\}}\,dt\,dz \\ &= \int\int f_{T,Z}(t,z)\mathbf{1}_{\{0 \le t \le w-z\}}\,dt\,dz \\ &= \int_0^w \int_0^{w-z} f_{T,Z}(t,z)\,dt\,dz. \end{aligned}$$

Since T and Z are independent, we know that

$$f_{T,Z}(t,z) = f_T(t) f_Z(z)$$

So, using the expression we obtained in class and the density of a unit uniform, we get

$$F_W(w) = \int_0^w \int_0^{w-z} f_T(t) f_Z(z)\, dt\, dz$$

Thus, since $Z \sim U(0,1)$, we get that

$$F_W(w) = \int_0^{1\wedge w} \int_0^{w-z} f_T(t)\, dt\, dz$$

For $w \in [0,1]$, we know that $w - z$ within the bounds of integration above is also in $[0,1]$. So,

$$F_W(w) = \int_0^w \int_0^{w-z} t\, dt\, dz = \int_0^w \tfrac{(w-z)^2}{2}\, dz = \tfrac{w^3}{6}\,.$$

As an alternative, we could have used a **geometric** argument in $3D$ similar to the one we used in class in $2D$.

For $w \in [1,2]$, we have that

$$\begin{aligned} F_W(w) &= \int_0^1 \int_0^{w-z} f_T(t)\, dt\, dz \\ &= \int_0^{w-1} \int_0^{w-z} f_T(t)\, dt\, dz + \int_{w-1}^1 \int_0^{w-z} f_T(t)\, dt\, dz \end{aligned} \tag{4.1}$$

mac

Let's focus on the first integral in the sum above. Due to the fact that f_T is piecewise defined, it's prudent to split the inner integral as follows

$$\begin{aligned} \int_0^{w-z} f_T(t)\, dt &= \int_0^1 f_T(t)\, dt + \int_1^{w-z} f_T(t)\, dt \\ &= \int_0^1 t\, dt + \int_1^{w-z} (2-t)\, dt \\ &= \frac{1}{2} + \int_{2-w+z}^1 u\, du \\ &= \frac{1}{2} + \frac{1}{2}\left(1 - (2-w+z)^2\right) \\ &= \frac{1}{2}\left(2 - (4 + w^2 + z^2 - 4w + 4z - 2wz)\right) \\ &= \frac{1}{2}\left(-2 - w^2 - z^2 + 4w - 4z + 2wz\right). \end{aligned}$$

Now, the entire first integral in (4.1) reads as

$$\begin{aligned}\int_0^{w-1}\int_0^{w-z} f_T(t)\,dt\,dz &= \frac{1}{2}\int_0^{w-1}\left(-2-w^2-z^2+4w-4z+2wz\right)\,dz\\ &= \frac{1}{2}\left(-2(w-1)-w^2(w-1)-\tfrac{(w-1)^3}{3}+4w(w-1)-4\left(\tfrac{(w-1)^2}{2}\right)+2w\left(\tfrac{(w-1)^2}{2}\right)\right)\\ &= \frac{1}{2}\left(-2w+2-w^3+w^2-\tfrac{1}{3}(w^3-3w^2+3w-1)+4w^2-4w-2(w^2-2w+1)+w(w^2-2w+1)\right)\\ &= \frac{1}{2}\left(-\tfrac{1}{3}w^3+2w^2-2w+\tfrac{1}{3}\right).\end{aligned}$$

The second integral in (4.1) is much easier. We get

$$\begin{aligned}\int_{w-1}^1\int_0^{w-z} f_T(t)\,dt\,dz &= \int_{w-1}^1\int_0^{w-z} t\,dt\,dz\\ &= \frac{1}{2}\int_{w-1}^1 (w-z)^2\,dz\\ &= \frac{1}{2}\int_{w-1}^1 u^2\,du\\ &= \frac{1}{6}(1-(w-1)^3)\\ &= \frac{1}{6}(1-(w^3-3w^2+3w-1))\\ &= \frac{1}{6}(2-w^3+3w^2-3w).\end{aligned}$$

Combining the two parts of (4.1), we obtain

$$F_W(w) = -\frac{1}{3}w^3+\frac{3}{2}w^2-\frac{3}{2}w+\frac{5}{6}.$$

Finally, for $w\in[2,3]$, we have that by symmetry with the first case

$$F_W(w) = \frac{(3-w)^3}{6}.$$

Upon differentiation, we get that

$$f_W(w) = \begin{cases}\frac{w^2}{2} & \text{for } w\in[0,1]\\ -w^2+3w-\frac{3}{2} & \text{for } w\in[1,2]\\ \frac{(3-w)^2}{2} & \text{for } w\in[2,3]\end{cases}$$

The density f_W is zero otherwise.

Method II: Convolution. Let us introduce the random variable $T=X+Y$, and find its density function. For every $a\in\mathbb{R}$

$$\begin{aligned}f_T(a) &= \int_{-\infty}^{+\infty} f_X(y)f_Y(a-y)\,dy\\ &= \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1]}(y)\mathbf{1}_{[0,1]}(a-y)\,dy.\end{aligned}$$

Now, we can conclude that $f_T(a) = 0$ for every $a \notin [0, 2]$. For all $a \in [0, 1]$, we have

$$f_T(a) = \int_0^a dy = a.$$

For all $a \in [1, 2]$, we have

$$f_T(a) = \int_{a-1}^1 dy = 1 - (a - 1) = 2 - a.$$

Written all in one expression:

$$f_T(a) = \begin{cases} a & \text{for } a \in [0, 1] \\ 2 - a & \text{for } a \in [1, 2] \\ 0 & \text{otherwise.} \end{cases}$$

Now, we can express the random variable W as $W = T + Z$. Since X, Y and Z were assumed independent, we can conclude that T and Z are independent as well. So, we can write the density of W as

$$f_W(a) = \int_{-\infty}^{+\infty} f_T(y) f_Z(a - y)\, dy$$

for all $a \in [0, 3]$. It is evident that f_W vanishes outside of this interval. It is convenient to partition the above integral into regions as follows:

$$\begin{aligned} f_W(a) &= \int_0^1 f_T(y) f_Z(a - y)\, dy + \int_1^2 f_T(y) f_Z(a - y)\, dy \\ &= \int_0^1 y \mathbf{1}_{[0,1]}(a - y)\, dy + \int_1^2 (2 - y) \mathbf{1}_{[0,1]}(a - y)\, dy. \end{aligned}$$

This representation leads us to consider different cases of the values of a separately. For $a \in [0, 1]$, we have

$$f_W(a) = \int_0^a y\, dy = \frac{1}{2}\, a^2 .$$

For $a \in [1, 2]$, we have

$$\begin{aligned} f_W(a) &= \int_{a-1}^1 y\, dy + \int_1^a (2 - y)\, dy \\ &= \frac{1}{2} \left[(1 - (a - 1)^2) - (2 - a)^2 + 1 \right] \\ &= \frac{1}{2} \left[2a - a^2 - 4 + 4a - a^2 + 1 \right] \\ &= \frac{1}{2} \left[-2a^2 + 6a - 3 \right]. \end{aligned}$$

For $a \in [2, 3]$, we have

$$f_W(a) = \int_{a-1}^2 (2 - y)\, dy = -\frac{1}{2} \left[0 - (3 - a)^2 \right] = \frac{1}{2}\, (a - 3)^2 .$$