MASD (Probability Part) Assignment 7

Hand-in in groups of 2 or 3 before November 4, 2021 at 10:00
One submission per group
Remember to include the names of all group members

Problem 1 (Random variables and their distributions):

1. Determine $\mathbb{E}|X|$ for $X \sim N(0,1)$.
2. Let X, Y be independent and uniformly distributed on the interval $[-1,1]$. Determine the PDF and CDF of $X+Y$ and $\min\{X,Y\}$.
3. Let X, Y be i.i.d. and uniformly distributed on $[0,1]$. Compute the variance of the random variable $3X+Y-1$.

Solution for Problem 1: For 1: We compute

$$\mathbb{E}|X| = \frac{1}{\sqrt{2\pi}}\int |x|\mathrm{e}^{-\frac{1}{2}x^2}\mathrm{d}x = \sqrt{\frac{2}{\pi}}\int_0^\infty x\,\mathrm{e}^{-\frac{1}{2}x^2}\mathrm{d}x = -\sqrt{\frac{2}{\pi}}\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{-\frac{1}{2}x^2}\mathrm{d}x = -\sqrt{\frac{2}{\pi}}\mathrm{e}^{-\frac{1}{2}x^2}\Big|_0^\infty = \sqrt{\frac{2}{\pi}}\,.$$

For 2: We start by considering $X+Y$. Since this is the sum of independent continuous random variables we can use the convolution rule to compute the PDF of $X+Y$. This is what we do first through

$$\begin{aligned} f_{X+Y}(z) &= \int_{\mathbb{R}} f_X(x) f_Y(z-x)\mathrm{d}x \\ &= \frac{1}{4}\int_{\mathbb{R}} \mathbb{1}_{[-1,1]}(x)\mathbb{1}_{[-1,1]}(z-x)\mathrm{d}x \\ &= \frac{1}{4}\mathrm{Vol}([-1,1]\cap[z-1,z+1])\,. \end{aligned}$$

Here $\mathrm{Vol}(\cdot)$ applied to the interval means its length. At this point we distinguish two cases. If $z \in [-2,0]$, then

$$[-1,1]\cap[z-1,z+1] = [-1,z+1]$$

and if $z \in [0,2]$, then

$$[-1,1]\cap[z-1,z+1] = [z-1,1]\,.$$

For all other cases $x \in \mathbb{R}\setminus[-2,2]$ the intersection is empty and $f_{X+Y}(z) = 0$. We conclude

$$f_{X+Y}(z) = \begin{cases} \frac{z+2}{4} & \text{if } z \in [-2,0] \\ \frac{2-z}{4} & \text{if } z \in [0,2] \end{cases} = \frac{2-|z|}{4}\mathbb{1}_{[-2,2]}(z)\,. \tag{0.1}$$

For good reason this is called the triangular distribution (Draw the function!).

Now we compute the CDF for $X+Y$. Since we already determined f_{X+Y} this is now simple via

$$\mathbb{P}(X+Y \le x) = \int_{-\infty}^x f_{X+Y}(y)\mathrm{d}y\,.$$

In particular, $F_{X+Y}(x) = 0$ for $x < -2$ and $F_{X+Y}(x) = 1$ for $x > 2$. Now let $x \in [-2,0]$. Then

$$F_{X+Y}(x) = \int_{-\infty}^x f_{X+Y}(y)\mathrm{d}y = \int_{-2}^x f_{X+Y}(y)\mathrm{d}y = \frac{1}{4}\int_{-2}^x (y+2)\mathrm{d}y = \frac{x^2-4}{8} + \frac{x+2}{2}\,.$$

For $x \in [0,2]$ we find

$$\begin{aligned}F_{X+Y}(x) &= \int_{-\infty}^{x} f_{X+Y}(y)\mathrm{d}y \\ &= \int_{-2}^{0} f_{X+Y}(y)\mathrm{d}y + \int_{0}^{x} f_{X+Y}(y)\mathrm{d}y \\ &\overset{(1)}{=} \frac{1}{2} + \frac{1}{4}\int_0^x (2-y)\mathrm{d}y \\ &= \frac{1}{2} + \frac{4x - x^2}{8}\,.\end{aligned}$$

In (1) we used $F_{X+Y}(0) = \frac{1}{2}$ which we already computed. As a sanity check we see that $F_{X+Y}(2) = 1$, which is good because X is a continuous random variable.

Now we turn our attention to $Z := \min\{X, Y\}$. Since we compute the minimum of two independent random variables, we start with the CDF and take the complement. We compute

$$\mathbb{P}(Z \le x) = 1 - \mathbb{P}(Z > x) = 1 - \mathbb{P}(X > x, Y > x) = 1 - \mathbb{P}(X > x)\mathbb{P}(Y > x) = 1 - (1 - F_X(x))^2\,.$$

Now let us compute F_X. Since X is uniformly distributed on $[-1,1]$ we have $F_X(x) = 0$ for $x < -1$ and $F_X(x) = 1$ for $x > 2$. For $x \in [-1,1]$ we find

$$F_X(x) = \frac{1}{2}\int_{-1}^{x} \mathrm{d}y = \frac{x+1}{2}\,.$$

We conclude that $F_Z(x) = 0$ for $x < -1$ and $F_Z(x) = 1$ for $x > 1$. In between we have for $x \in [-1,1]$ the formula

$$F_Z(x) = 1 - \left(1 - \frac{x+1}{2}\right)^2 = 1 - \frac{(x-1)^2}{4}\,.$$

To determine the PDF $f_Z(x) = F_Z'(x)$ we differentiate and find

$$f_Z(x) = \frac{1-x}{2}\mathbb{1}_{[-1,1]}(x)\,.$$

For 3: Since adding a constant does not change the variance, we have $\operatorname{Var}(3X+Y-1) = \operatorname{Var}(3X+Y)$. By independence of X and Y we get $\operatorname{Var}(3X+Y) = \operatorname{Var}(3X) + \operatorname{Var} Y = 9\operatorname{Var} X + \operatorname{Var} Y$. Since X and Y are both uniformly distributed on $[0,1]$, we find

$$\operatorname{Var}(3X+Y-1) = 10 \operatorname{Var} X = \frac{10}{12} = \frac{5}{6}\,.$$

Problem 2 (Joint distribution):

1. Let X, Y be i.i.d. uniformly distributed on $\{-1, 0, 1\}$. Compute the covariance $\operatorname{Cov}(X, X-Y)$. What is the joint PMF for the joint distribution of X and $X - Y$?

2. Compute $\mathbb{E}e^{\frac{1}{2}(X^2-Y^2)}$, where X, Y are jointly Gaussian with $(X,Y) \sim N(0, A)$ and
$$A = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\,.$$

3. Let X and Y be jointly continuous random variables with joint PDF
$$f_{X,Y}(x,y) = \frac{6}{7}(x+y)^2\mathbb{1}_{[0,1]}(x)\,\mathbb{1}_{[0,1]}(y)\,.$$
Determine the PDF of the distribution of X.

Solution for Problem 2: For 1: To compute the covariance we realise that $\mathrm{Cov}(X, X-Y) = \mathrm{Var}\,X - \mathrm{Cov}(X,Y) = \mathrm{Var}\,X$ because X and Y are independent. The variance of X is computed via

$$\mathrm{Var}\,X = \mathbb{E}X^2 = \frac{1}{3}((-1)^2 + 0^2 + 1^2) = \frac{2}{3}\,.$$

For the joint PMF of X and $X - Y$ we compute

$$\begin{aligned} P_{X,X-Y}(x,z) &= \mathbb{P}(X = x, X - Y = z) \\ &= \mathbb{P}(X = x, Y = x - z) \\ &= P_X(x)P_Y(x-z) \\ &= \frac{1}{9}\mathbb{1}_{\{-1,0,1\}}(x)\mathbb{1}_{\{-1,0,1\}}(x-z)\,. \end{aligned}$$

This is the PMF of the uniform distribution on $\{(x,z) \in \mathbb{Z}^2 : |x| \le 1, |x - z| \le 1\}$.

For 2: The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \frac{\sqrt{3}}{2\pi}\mathrm{e}^{-x^2-y^2+xy}\,.$$

Therefore we have

$$\begin{aligned} \mathbb{E}\mathrm{e}^{\frac{1}{2}(X^2-Y^2)} &= \frac{\sqrt{3}}{2\pi}\int \mathrm{e}^{\frac{1}{2}x^2-\frac{1}{2}y^2}\mathrm{e}^{-x^2-y^2+xy}\mathrm{d}x\mathrm{d}y \\ &= \frac{\sqrt{3}}{2\pi}\int \mathrm{e}^{-y^2}\int \mathrm{e}^{-\frac{1}{2}(x-y)^2}\mathrm{d}x\mathrm{d}y \\ &= \frac{\sqrt{3}}{\sqrt{2\pi}}\int \mathrm{e}^{-y^2}\mathrm{d}y = \sqrt{\frac{3}{2}}\,. \end{aligned}$$

For 3: We use that the PDF of the marginal distribution (distribution of X) results from integrating over the variable corresponding to Y, i.e.

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\mathrm{d}y = \frac{6}{7}\mathbb{1}_{[0,1]}(x)\int_0^1 (x+y)^2\,\mathrm{d}y = \frac{6}{7}\mathbb{1}_{[0,1]}(x)\int_x^{1+x} y^2\,\mathrm{d}y = \frac{2}{7}(3x^2+3x+1)\mathbb{1}_{[0,1]}(x)\,.$$

Thus, X is a continuous random variable with PDF as above.

Problem 3 (Statistics): We consider the statistical model $(\mathbb{P}_\theta)_{\theta>0}$ with $\mathbb{P}_\theta = \mathrm{Exp}(\theta)$ and are given a sample $X_1, \dots, X_n$ of n independent observations distributed as $\mathbb{P}_\theta$. Determine the MLE $T(x)$ for this model.

Solution for Problem 3: To determine the MLE we need to maximise the function

$$\prod_{i=1}^n \rho_\theta(x_i) = \prod_{i=1}^n (\theta\,\mathrm{e}^{-\theta x_i}) = \theta^n \mathrm{e}^{-\theta\sum_{i=1}^n x_i} = \mathrm{e}^{n\log\theta - \theta\sum_{i=1}^n x_i}$$

in θ for any given sample $x_1, \dots, x_n > 0$. The maximum of this expression is precisely the maximum of the exponent $h(\theta) = n\log\theta - \theta\sum_{i=1}^n x_i$ in θ, i.e.

$$T(x) = \mathrm{argmax}_{\theta>0}\left(n\log\theta - \theta\sum_{i=1}^n x_i\right)\,.$$

To determine the maximum we differentiate in θ and find

$$h'(\theta) = \frac{n}{\theta} - \sum_{i=1}^n x_i\,.$$

Thus $T(x) = \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^{-1}$.