DPT Exercise 6

github.com/aiulus/dpt-ss25

START: 12:15

Exercise 1. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $X, Y : \Omega \to \mathbb{R}$ two $\mathbb{R}$-valued RVs with joint density

$$p_{X,Y}(x,y) = \exp(-y)\chi_A(x,y) \quad \text{for } A = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le y\} .$$

We can also write this equivalently as

$$p_{X,Y}(x,y) = \begin{cases} \exp(-y) & \text{if } x, y \in \mathbb{R} \text{ and } 0 \le x \le y , \\ 0 & \text{otherwise} . \end{cases}$$

(i) Verify that $p_{X,Y}(x,y)$ is a valid pdf.

(ii) Compute the conditional expectation $\mathbb{E}[X \mid Y]$.

(iii) Compute the conditional expectation $\mathbb{E}[Y \mid X]$.

Exercise 2. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $X, Y : \Omega \to \mathbb{R}$ two independent RVs with densities $p_X, p_Y : \mathbb{R} \to \mathbb{R}_{\ge 0}$.

(i) Let $Z := X + Y$. Then compute the pdf of Z and show that it is given by

$$p_Z(z) = \int_{\mathbb{R}} p_X(z-y)\, p_Y(y)\, dy .$$

Remark: Generally, for two functions $f, g : \mathbb{R} \to \mathbb{R}$, we call $(f \star g)(t) := \int_{-\infty}^{\infty} f(\tau)g(t-\tau)\, d\tau$ the *convolution* of f and g. (Note that this is symmetric.) Hence, $p_Z(z)$, the density of the sum of X and Y, is given by the *convolution* of the densities $p_X(x)$ and $p_Y(y)$ of X and Y. Accordingly, one sometimes also calls Z the convolution of X and Y (even though it's just the sum of X and Y and we didn't really need an additional name).
Note: This also holds more generally for independent $\mathbb{R}^n$-valued random variables.
[Hint: Consider the set $A_z := \{(x,y) \in \mathbb{R}^n \mid x + y \le z\}$ and compute $F_Z(z) = P(Z|lez)$ using the independence of X, Y. Finally, remember that by definition $F_Z(z) = \int_{-\infty}^{z} p_Z(z)\, dz$.]

(ii) Let $X, Y \sim \mathrm{Exp}(\lambda)$ be independent with $\lambda > 0$. Calculate the density of $X + Y$.

(iii) Let $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ be independent with $\mu_x, \mu_y \in \mathbb{R}$ and $\sigma_x, \sigma_y > 0$. Show that

$$X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2) .$$

This is a fairly long and tedious computation, which would be too lengthy and error prone for the exam, but is a good exercise to get a feeling for how computations with Gaussians work. Start from the convolution on one side and the final result—the density of $\mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$—and try to "match them from both directions."

Exercise 3. Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a random variable on Ω.

(i) Let $X : \Omega \to \mathbb{N}$ be a discrete random variable. Show that

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} P(X \ge n)$$

and describe in words in a single sentence what that means.

(ii) Assume now that $X : \Omega \to \mathbb{N}$ is geometrically distributed with parameter $p \in (0,1)$, $X \sim \mathrm{Geo}(p)$. Use (i) to calculate the expectation $\mathbb{E}[X]$.
[Hint: Recall the partial sum formula of the geometric series.]

(iii) If $X : \Omega \to \mathbb{R}_{\geq 0}$ is a non-negative, continuous RVRV with cdf F_X, then it holds that

$$\mathbb{E}[X] = \int_0^\infty 1 - F_X(x)\, dx \,.$$

Using this, calculate the expectation of a $X \sim \mathrm{Exp}(\lambda)$, where $\lambda > 0$.

Exercise 4 (application + coding)**.** Consider the following (oversimplified) setting: We are at the conveyor-belt of a fishing trawler and should distinguish between different types of fish. We want to separate trouts ("Forelle"), which is the "on average smaller fish") from salmon ("Lachs"), the "on average larger fish." We label trouts as $Y = 0$ and salmon as $Y = 1$. We distinguish them by size, which we denote by X. We also call X a "feature" here. We assume that we assess size with some value in $[0, 1]$, where 0 is smallest and 1 is largest. Mathematically, this can be modeled in the following way. Let X and Y be $[0, 1]$-valued and $\{0, 1\}$-valued random variables on the same probability space, respectively. We assume that there are equally many trouts and salmons in the water, which we model by the prior probabilities

$$P(Y = 0) = P(Y = 1) = \frac{1}{2} \,.$$

Moreover, from experience we know how the sizes of trouts and salmons are distributed. We assume this to be given by the likelihood densities

$$p(x \mid Y = 0) = \begin{cases} 2 - 2x & \text{if } x \in [0, 1] \,, \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad p(x \mid Y = 1) = \begin{cases} 2x & \text{if } x \in [0, 1] \,, \\ 0 & \text{otherwise} \,. \end{cases}$$

We want to use inverse transform sampling to sample X for each class $Y \in \{0, 1\}$. We proceed step by step.

(i) Work out the required inverse transformation formulas "on paper".

(ii) Implement a function `generate_data(N: int)` that, provided and integer `N`, generates `N` realizations of (X, Y) and returns these observations in an array of shape $(N, 2)$. Then plot the the empirical cumulative distribution functions (for example, `scipy.stats.ecdf`) and demonstrate that it coincides with the theoretical result from (i).

(iii) To classify $Y = 0$ (trout) and $Y = 1$ (salmon), consider the following threshold-based classifiers with threshold $x_t \in [0, 1]$.

- Classifier A: For a fish with given size x, we predict it belongs to class

$$\hat{y} := f_{x_t}^{(A)}(x) := \begin{cases} 0 & \text{if } x \leq x_t \,, \\ 1 & \text{if } x > x_t \,. \end{cases}$$

- Classifier B ("Reverse of A"): For a fish with given size x, we predict it belongs to class

$$\hat{y} := f_{x_t}^{(B)}(x) := \begin{cases} 1 & \text{if } x \leq x_t \,, \\ 0 & \text{if } x > x_t \,. \end{cases}$$

- Classifier C ("Guessing"): For any fish, we simply flip a coin, that is,

$$\hat{y} := f_{x_t}^{(C)}(x) := \begin{cases} 0 & \text{with prob. } \frac{1}{2} \,, \\ 1 & \text{otherwise} \,. \end{cases}$$

  In other words, $\hat{y} \sim \mathrm{Ber}(\frac{1}{2})$.

- Classifier D ("Everything's a salmon"): We're lazy and simply throw all fish into the salmon bucket, that is,
$$\hat{y} := f_{x_t}^{(D)}(x) := 1\,.$$

Calculate the *average error rate* for each of the classifiers in terms of $x_t \in [0,1]$ given by

$$\mathbb{E}_{X,Y}\left[\chi_{\{f_{x_t}(X)\neq Y\}}\right] = \iint \chi_{\{f_{x_t}(x)\neq y\}}(x,y)\, p(x \mid y)\, p(y)\, dx\, dy$$

manually "on paper".

(iv) Confirm experimentally for $x_t \in \{0.05, 0.1, 0.15, \ldots, 0.9, 0.95\}$ that the computed error rates match empirical values (using a large number of samples generated as in (ii)). Repeat each of these tests with 10 different test datasets of the same size N (using different random seeds) and compute the mean and standard deviation of the estimated average error rates. Use the sample sizes $N \in \{1000, 10000\}$. How does the standard deviation of the estimated error rate decrease with increasing N?

Exercise 5 (Bonus). Use inverse transform sampling to sample from a 2-dimensional Gaussian with

$$\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

More precisely:

(i) Derive the transformation to sample from a one dimensional Gaussian $\mathcal{N}(0,1)$.

(ii) **Coding:** Use the previous calculated transformation to sample from $\mathcal{N}(\mu,\Sigma)$. You can use the fact that $A\boldsymbol{Z} + b \sim \mathcal{N}(A\mu + b, A\Sigma A^T)$ for $A \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$ where $\boldsymbol{Z} \sim \mathcal{N}(\mu,\Sigma)$. (You can assume that $A\Sigma A^T \succ 0$.)
[Hint: Use the Cholesky decomposition.]

Exercise 1. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $X, Y : \Omega \to \mathbb{R}$ two $\mathbb{R}$-valued RVs with joint density

$$p_{X,Y}(x,y) = \exp(-y)\underbrace{\chi_A(x,y)}_{= \mathbb{1}\{(x,y) \in A\}} \quad \text{for } A = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le y\}\,.$$

We can also write this equivalently as

$$p_{X,Y}(x,y) = \begin{cases} \exp(-y) & \text{if } x, y \in \mathbb{R} \text{ and } 0 \le x \le y\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

(i) Verify that $p_{X,Y}(x,y)$ is a valid pdf.

$$\int_{(x,y)\in\mathbb{R}^2} p_{X,Y}(x,y)\, d(x,y) = \int_{(x,y)\in\mathbb{R}^2} e^{-y} \cdot \underline{\chi_A(x,y)}\, d(x,y)$$

$$= \int_{(x,y)\in\mathbb{R}^2} e^{-y} \cdot \mathbb{1}\{0 \le x \le y\}\, d(x,y)$$

$$= \int_0^\infty \int_x^\infty e^{-y}\, dy\, dx = \int_0^\infty \left[-e^{-y}\right]_x^\infty dx$$

$$= \int_0^\infty \left(-\lim_{y\to\infty} e^{-y} - (-e^{-x})\right) dx$$

$$= \int_0^\infty e^{-x}\, dx = \left[-e^{-x}\right]_0^\infty = 0 - (-e^0) = 1 \quad \blacksquare$$

$A = \{(x,y) \mid 0 \le x \le y\}$

$A_x = \{y \in \mathbb{R} \mid y \ge x \,;\, x \ge 0 \text{ fixed } x \in \mathbb{R}_{\ge 0}\}$

$A_y = \{x \in \mathbb{R} \mid x \le y \text{ for fixed } y \in \mathbb{R}_{\ge 0}\}$

1. $y \in A_x \wedge x \in \mathbb{R}_{\ge 0}$

2. $x \in A_y \wedge y \in \mathbb{R}_{\ge 0}$

1. $\displaystyle\int_{y\in A_x} \int_{x=0}^{\infty} \square \, dx\, dy$

2. $\displaystyle\int_{y=0}^{\infty} \int_{x\in A_y} \square \, dx\, dy$

Exercise 1. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $X, Y : \Omega \to \mathbb{R}$ two $\mathbb{R}$-valued RVs with joint density

$$p_{X,Y}(x,y) = \exp(-y)\chi_A(x,y) \quad \text{for } A = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le y\}\,.$$

We can also write this equivalently as

$$p_{X,Y}(x,y) = \begin{cases} \exp(-y) & \text{if } x, y \in \mathbb{R} \text{ and } 0 \le x \le y\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

(ii) Compute the conditional expectation $\mathbb{E}[X \mid Y]$.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot p_X(x)\, dx$$

(ii) $$\mathbb{E}[X \mid Y] = \int_{-\infty}^{\infty} x \cdot p_{X|Y=y}(x)\, dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{y} \cdot \mathbb{1}\{0 \le x \le y\}\, dx = \frac{1}{y} \cdot \int_0^y x \; \chi_{A_{\geq 0}}(y)\, dx$$

$$= \frac{1}{y} \cdot \chi_{A_{\geq}}(y) \cdot \int_0^y x\, dx = \frac{1}{y} \cdot \chi_{A_{\geq 0}}(y) \cdot \left[\frac{1}{2}x^2\right]_0^y$$

$$= \frac{1}{y} \cdot y^2/2 \cdot \chi_{A_{\geq 0}}(y) = y/2 \cdot \chi_{A_{\geq 0}}(y) = \begin{cases} y/2 & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

» Side computations:

$$p_{X|Y=y}(x) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$ » Marginal density

$$\Rightarrow p_Y(y) = \int_{-\infty}^{\infty} e^{-y} \cdot \underbrace{\mathbb{1}\{0 \le x \le y\}}_{= \chi_A(x,y)}\, dx = \int_0^y e^{-y}\, dx = e^{-y} \cdot \int_0^y 1\, dx = e^{-y} \cdot y$$

$$= \frac{\cancel{e^{-y}} \cdot \chi_A(x,y)}{\cancel{e^{-y}} \cdot y} = \frac{1}{y} \cdot \chi_A(x,y)$$

Exercise 1. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $X, Y : \Omega \to \mathbb{R}$ two $\mathbb{R}$-valued RVs with joint density

$$p_{X,Y}(x,y) = \exp(-y)\chi_A(x,y) \quad \text{for } A = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le y\}\,.$$

We can also write this equivalently as

$$p_{X,Y}(x,y) = \begin{cases} \exp(-y) & \text{if } x, y \in \mathbb{R} \text{ and } 0 \le x \le y\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

(iii) Compute the conditional expectation $\mathbb{E}[Y \mid X]$.

$$\mathbb{E}[Y|X] = \int_{-\infty}^{\infty} y \cdot p_{Y|X=x}(y)\, dy \overset{(2)}{=} \int_{-\infty}^{\infty} y \cdot e^{x-y} \cdot \mathbb{1}\{0 \le x \le y\}\, dy = e^{x} \cdot \chi_{\mathbb{R}_{\ge 0}}(x) \cdot \int_{x}^{\infty} y \cdot e^{-y}\, dy$$

$$\overset{(*)}{=} e^{x} \cdot \chi_{\mathbb{R}_{\ge 0}}(x) \cdot \left[-e^{-y} \cdot (y+1)\right]_x^{\infty} = e^{x} \cdot \chi_{\mathbb{R}_{\ge 0}}(x) \cdot e^{-x} \cdot (x+1)$$

$$= (x+1) \cdot \chi_{\mathbb{R}_{\ge 0}}(x)$$

$$= \begin{cases} x+1 & \text{if } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

$$(1)\ p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\, dy$$

$$= \int_{-\infty}^{\infty} e^{-y} \cdot \mathbb{1}\{0 \le x \le y\}\, dy$$

$$= \chi_{\mathbb{R}_{\ge 0}}(x) \cdot \int_{x}^{\infty} e^{-y}\, dy$$

$$= \chi_{\mathbb{R}_{\ge 0}}(x) \cdot \left[-e^{-y}\right]_x^{\infty}$$

$$= \chi_{\mathbb{R}_{\ge 0}}(x) \cdot e^{-x}$$

$$(2)\ p_{Y|X=x}(y) = \frac{p_{X,Y}(x,y)}{p_X(y)} = \frac{e^{-y} \cdot \mathbb{1}\{\overbrace{0 \le x}\le y\}}{e^{-x} \cdot \mathbb{1}\{\cancel{0 \le} x\}}$$

$$= e^{x-y}\, \chi_A(x,y)$$

$(*)$ $\int f'g = fg - \int fg'$ or Wolfram Alpha

Exercise 2. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $X, Y : \Omega \to \mathbb{R}$ two independent RVs with densities $p_X, p_Y : \mathbb{R} \to \mathbb{R}_{\geq 0}$.

(i) Let $Z := X + Y$. Then compute the pdf of Z and show that it is given by

$$p_Z(z) = \underline{\int_{\mathbb{R}} p_X(z-y)\, p_Y(y)\, dy} \, .$$

Remark: Generally, for two functions $f, g : \mathbb{R} \to \mathbb{R}$, we call $(f \star g)(t) := \int_{-\infty}^{\infty} f(\tau) g(t-\tau)\, d\tau$ the *convolution* of f and g. (Note that this is symmetric.) Hence, $p_Z(z)$, the density of the sum of X and Y, is given by the *convolution* of the densities $p_X(x)$ and $p_Y(y)$ of X and Y. Accordingly, one sometimes also calls Z the convolution of X and Y (even though it's just the sum of X and Y and we didn't really need an additional name).

Note: This also holds more generally for independent $\mathbb{R}^n$-valued random variables.

[Hint: Consider the set $A_z := \{(x, y) \in \mathbb{R}^n \mid x + y \leq z\}$ and compute $F_Z(z) = P(Z|lez)$ using the independence of X, Y. Finally, remember that by definition $F_Z(z) = \int_{-\infty}^{z} p_Z(z)\, dz$.]

The Law of total probability

$$p_z(z) = P(Z=z) = P(X+Y=z) = \int_{x\in\mathbb{R}} \underbrace{P(X+Y=z \mid X=x)} \cdot \underbrace{P(X=x)}_{p_X(x)} \, dx$$

$= P(X+Y=z \mid X=x)$

$= P(Y=z-x \mid X=x)$

$= P_{Y|X=x}(z-x)$ $X \perp\!\!\!\perp Y$

$= p_Y(z-x)$

by def; $X \perp\!\!\!\perp Y$ (both cont. RV's)

$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$

$\Leftrightarrow p_{X|Y}(x) = p_X(x)$, $p_{Y|X}(y) = p_Y(y)$

$$= \int_{x\in\mathbb{R}} p_Y(z-x) \cdot p_X(x) \, dx$$

$$= \int_{y\in\mathbb{R}} p_X(z-y) \cdot p_Y(y) \, dy$$

(ii) Let $X, Y \sim \mathrm{Exp}(\lambda)$ be independent with $\lambda > 0$. Calculate the density of $X + Y$.

(iii) Let $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ be independent with $\mu_x, \mu_y \in \mathbb{R}$ and $\sigma_x, \sigma_y > 0$. Show that

$$X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)\,.$$

This is a fairly long and tedious computation, which would be too lengthy and error prone for the exam, but is a good exercise to get a feeling for how computations with Gaussians work. Start from the convolution on one side and the final result—the density of $\mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$—and try to "match them from both directions."

Convolution formula ($X \perp\!\!\!\perp Y$!)

$$p_{X+Y}(z) = \int_{y\in\mathbb{R}} p_X(z-y)\cdot p_Y(y)\, dy$$

$$p_{\mathrm{Exp}(\lambda)}(x) = \lambda\cdot e^{-\lambda x}\cdot \chi_{\mathbb{R}_{\geq 0}}(x) \Rightarrow \begin{cases} p_X(x) = \lambda\cdot e^{-\lambda x}\cdot \chi_{\mathbb{R}_{\geq 0}}(x) \\ p_Y(y) = \lambda\cdot e^{-\lambda y}\cdot \chi_{\mathbb{R}_{\geq 0}}(y) \end{cases}$$

$$\chi_A(x)\cdot \chi_B(x) = \chi_{A\cap B}(x) \quad (*)$$

$$p_z(z) = \int_{-\infty}^{\infty} \lambda\, e^{-\lambda\cdot(z-y)}\cdot \chi_{\mathbb{R}_{\geq 0}}(z-y)\cdot \lambda\cdot e^{-\lambda y}\cdot \chi_{\mathbb{R}_{\geq 0}}(y)\, dy$$

$$= \lambda^2\cdot \int_{-\infty}^{\infty} e^{-\lambda(z-y+y)}\cdot \underbrace{\mathbb{1}\{z-y\geq 0\}\cdot \mathbb{1}\{y\geq 0\}}\, dy$$

$$\overset{(*)}{=} \mathbb{1}\{z\geq y \wedge y\geq 0\} = \mathbb{1}\{z\geq y\geq 0\} = \mathbb{1}\{y\in[0,z]\}$$

$$= \lambda^2\cdot e^{-\lambda z}\cdot \chi_{\mathbb{R}_{\geq 0}}(z)\cdot \int_0^z 1\, dy = \lambda^2\cdot e^{-\lambda z}\cdot z\cdot \chi_{\mathbb{R}_{\geq 0}}(z) = \begin{cases} \lambda^2 z\, e^{-\lambda z} & \text{if } z\geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Exercise 3. Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a random variable on Ω.

(i) Let $X : \Omega \to \mathbb{N}$ be a discrete random variable. Show that

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} P(X \geq n)$$

and describe in words in a single sentence what that means.

(ii) Assume now that $X : \Omega \to \mathbb{N}$ is geometrically distributed with parameter $p \in (0,1)$, $X \sim \mathrm{Geo}(p)$. Use (i) to calculate the expectation $\mathbb{E}[X]$.
[Hint: Recall the partial sum formula of the geometric series.]

(i) $\mathbb{E}[X] = \sum_{n=0}^{\infty} n \cdot P(X=n)$ Trick: $n = \sum_{k=1}^{n} 1$

$= \sum_{n=0}^{\infty} \sum_{k=1}^{n} P(X=n) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \mathbb{1}\{1 \leq k \leq n\} \cdot P(X=n) = \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \mathbb{1}\{1 \leq k \leq n\} \cdot P(X=n)$

$\mathbb{1}\{1 \leq k \leq n\} = \mathbb{1}\{k \in [n]\} = \chi_{[n]}(k)$

$= \sum_{k=0}^{n} \mathbb{1}\{k \geq 1\} \sum_{n=k}^{\infty} P(X=n)$

$= \sum_{k=1}^{\infty} P(X \geq k)$

$P(X=0) = 0$? or $\mathbb{N} \setminus \{0\}$?

Cheat sheet!

(ii) $P_{Geo(p)}(x) = p \cdot (1-p)^{x-1}$ (for $x \in \mathbb{N}_{\geq 1}$) ; GEOMETRIC SERIES ; $\sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q}$, $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$ (for $q \in (0,1)$)

$\mathbb{E}[X] = \sum_{n=0}^{\infty} P(X \geq n) = \sum_{n=0}^{\infty} 1 - P(X \leq n-1)$;

$= \sum_{n=0}^{\infty} 1 - (1 - (1-p)^{n-1})$ Index shift + $P(X=0)=0$

$= \sum_{n=0}^{\infty} (1-p)^{n-1} = \sum_{n=0}^{\infty} (1-p)^{n}$

$= \frac{1}{1-(1-p)} = 1/p$

$P(X \leq n-1) = \sum_{k=1}^{n-1} p \cdot (1-p)^{k-1}$

$= p \cdot \sum_{k=1}^{n-1} (1-p)^{k-1}$ Index shift

$= p \cdot \sum_{k=0}^{n-2} (1-p)^{k} = p \cdot \frac{1-(1-p)^{n-1}}{1-(1-p)}$

$= 1 - (1-p)^{n-1}$

(iii) If $X : \Omega \to \mathbb{R}_{\geq 0}$ is a non-negative, continuous RVRV with cdf F_X, then it holds that

$$\mathbb{E}[X] = \int_0^\infty 1 - F_X(x)\, dx\,.$$ proof analogous to (1)

Using this, calculate the expectation of a $X \sim \mathrm{Exp}(\lambda)$, where $\lambda > 0$.

$F_X(x) = (1 - e^{-\lambda x}) \cdot \chi_{\mathbb{R}_{\geq 0}}(x)$

$\Rightarrow \mathbb{E}[X] = \int_0^\infty 1 - (1 - e^{-\lambda x})\, dx = \int_0^\infty e^{-\lambda x}\, dx = \left[-\frac{1}{\lambda} \cdot e^{-\lambda x} \right]_0^\infty$

$= 0 - \left(-\frac{1}{\lambda} \cdot e^{-\lambda 0} \right) = 1/\lambda$

Exercise 4 (application + coding)**.** Consider the following (oversimplified) setting: We are at the conveyor-belt of a fishing trawler and should distinguish between different types of fish. We want to separate trouts ("Forelle"), which is the "on average smaller fish") from salmon ("Lachs"), the "on average larger fish." We label trouts as $Y = 0$ and salmon as $Y = 1$. We distinguish them by size, which we denote by X. We also call X a "feature" here. We assume that we assess size with some value in $[0, 1]$, where 0 is smallest and 1 is largest. Mathematically, this can be modeled in the following way. Let X and Y be $[0, 1]$-valued and $\{0, 1\}$-valued random variables on the same probability space, respectively. We assume that there are equally many trouts and salmons in the water, which we model by the prior probabilities

$$P(Y = 0) = P(Y = 1) = \frac{1}{2}\,.$$

Moreover, from experience we know how the sizes of trouts and salmons are distributed. We assume this to be given by the likelihood densities

$$p(x \mid Y = 0) = \begin{cases} 2 - 2x & \text{if } x \in [0, 1]\,, \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad p(x \mid Y = 1) = \begin{cases} 2x & \text{if } x \in [0, 1]\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

$u \sim \mathrm{Unif}[0,1]$

$(*)$

We want to use inverse transform sampling to sample X for each class $Y \in \{0, 1\}$. We proceed step by step.

(1) $Y = 0$: $F_{X|Y=0}(x) = (2x - x^2) \cdot \cancel{\chi_{[0,1]}}(x) := u$

$\Leftrightarrow -(x-1)^2 + 1 = u \Leftrightarrow x = 1 + \sqrt{1-u}$

$Y = 1$: $F_{X|Y=1}(x) = x^2 \cdot \cancel{\chi_{[0,1]}}(x) := u \Leftrightarrow x =$

$2x - x^2 = u$

$-(x^2 - 2x + 1) + 1 = u$

$-(x-1)^2 + 1 = u$

$(x-1)^2 = 1 - u$

(i) Work out the required inverse transformation formulas “on paper”.

(ii) Implement a function `generate_data(N: int)` that, provided and integer `N`, generates `N` realizations of (X, Y) and returns these observations in an array of shape $(N, 2)$. Then plot the the empirical cumulative distribution functions (for example, `scipy.stats.ecdf`) and demonstrate that it coincides with the theoretical result from (i).

(iii) To classify $Y = 0$ (trout) and $Y = 1$ (salmon), consider the following threshold-based classifiers with threshold $x_t \in [0, 1]$.

- Classifier A: For a fish with given size x, we predict it belongs to class

$$\hat{y} := f_{x_t}^{(A)}(x) := \begin{cases} 0 & \text{if } x \le x_t\,, \\ 1 & \text{if } x > x_t\,. \end{cases}$$

- Classifier B ("Reverse of A"): For a fish with given size x, we predict it belongs to class

$$\hat{y} := f_{x_t}^{(B)}(x) := \begin{cases} 1 & \text{if } x \le x_t\,, \\ 0 & \text{if } x > x_t\,. \end{cases}$$

- Classifier C ("Guessing"): For any fish, we simply flip a coin, that is,

$$\hat{y} := f_{x_t}^{(C)}(x) := \begin{cases} 0 & \text{with prob. } \frac{1}{2}\,, \\ 1 & \text{otherwise}\,. \end{cases}$$

  In other words, $\hat{y} \sim \mathrm{Ber}(\frac{1}{2})$.

- Classifier D ("Everything's a salmon"): We're lazy and simply throw all fish into the salmon bucket, that is,

$$\hat{y} := f_{x_t}^{(D)}(x) := 1\,.$$

Calculate the *average error rate* for each of the classifiers in terms of $x_t \in [0, 1]$ given by

$$\mathbb{E}_{X,Y}\left[\chi_{\{f_{x_t}(X) \neq Y\}}\right] = \iint \chi_{\{f_{x_t}(x) \neq y\}}(x, y)\, p(x \mid y)\, p(y)\, dx\, dy$$

manually "on paper".

(A) $\mathbb{E}_{X,Y}\left[\mathbb{1}\{f_{x_t}^A(X) \neq Y\}\right] = \iint \underline{\mathbb{1}\{f_{x_t}^A(x) \neq y\}} \cdot p(x|y) \cdot p(y)\, dx\, dy$

$$= \int_{x \in \mathbb{R}} \sum_{y \in \{0,1\}} \mathbb{1}\{f_{x_t}^A(x) \neq y\} \cdot p(x|y) \cdot p(y)\, dx$$

$$= \int_{x\in\mathbb{R}} \underbrace{\mathbb{1}\{f_{x_t}^A(x) = 1\}}_{\Leftrightarrow\ x > x_t} \cdot \underbrace{p(x|Y=0)}_{2(1-x)\cdot\chi_{[0,1]}(x)} \cdot \underbrace{p_Y(0)}_{1/2}\, dx + \int_{x\in\mathbb{R}} \underbrace{\mathbb{1}\{f_{x_t}^A(x) = 0\}}_{x \le x_t} \cdot \underbrace{p(x|Y=1)}_{2x\cdot\chi_{[0,1]}(x)} \cdot \underbrace{p_Y(1)}_{1/2}\, dx$$

» $x_t \in (0,1)$

$$= \int_{x_t}^{1} 1 - x\, dx + \int_{0}^{x_t} x\, dx = \left[x - \frac{1}{2}x^2\right]_{x_t}^{1} + \left[\frac{1}{2}x^2\right]_0^{x_t} \quad ↯$$

(iv) Confirm experimentally for $x_t \in \{0.05, 0.1, 0.15, \ldots, 0.9, 0.95\}$ that the computed error rates match empirical values (using a large number of samples generated as in (ii)). Repeat each of these tests with 10 different test datasets of the same size N (using different random seeds) and compute the mean and standard deviation of the estimated average error rates. Use the sample sizes $N \in \{1000, 10000\}$. How does the standard deviation of the estimated error rate decrease with increasing N?