TTK4250 Sensor Fusion
Assignment 1

Hand in: *Friday 6. September 23:59* on Blackboard.

This assignment should be handed in on Blackboard, as a single PDF file, before the deadline. You are supposed to show how you got to each answer unless told otherwise. If you struggle, we encourage you to ask for help from a classmate or come to the exercise class on Monday.

Task 1: *The CDF as a random variable*

Let X be a continuous-valued random variable with cumulative distribution function $P_X(x) = \Pr(X \leq x)$. Show that the random variable $Y = P_X(X)$ is uniformly distributed over $[0, 1]$.

You can assume that P_X is differentiable and has an inverse for $Y \in [0, 1]$, i.e. if $y = P_X(x)$ then $x = P_X^{-1}(y)$ for valid x and y.

> *Hint:* The random variable itself, and not a fixed realization, is taken as an argument to its CDF. Keep in mind that $P_X(x)$ is a monotonous function, and that $P_X\left(P_X^{-1}(y)\right) = y$ by definition of the inverse function.
>
> Approach 1: Expand the cdf of Y, $P_Y(y)$ into its probability definition (see Definition 2.2.1). Using the definition of Y and invertibility of P_X, you should be able to rewrite this in terms of the definition of $P_X(x)$ (see Definition 2.2.1) with x as a function of y.
>
> Approach 2: Use transform of random variables. Since P_X is simply a function, $P_X(X)$ is simply a transform of a random variable. In this case you need differentiability, and potentially the calculus result
>
> $\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} = \left(\frac{\mathrm{d}g(x)}{\mathrm{d}x}\Big|_{x=g^{-1}(y)}\right)^{-1}$.

Task 2: *Some results regarding the Poisson distribution*

(a) Let N be a Poisson distributed random variable with parameter λ. Show that its probability generating function is $e^{\lambda(t-1)}$.

> *Hint:* The exponential function has the Taylor expansion $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$.

(b) Show that the probability generating function of a Binomial distributed Random variable M, with probability parameter r and number parameter n, is $(1 - r + rt)^n$.

> *Hint:* You can use PGFs with the fact that the Binomial distribution is a sum of n i.i.d. Bernoulli distributions (see Example 2.7 — Sum of Bernoulli and Poisson, and Exercise 2.5 for help) or you can use the Binomial theorem, which states that $\sum_{i=0}^{n} \frac{n!}{i!(n-i)!} a^i b^{(n-i)} = (a + b)^n$ along with the definitions of PGF and Binomial distribution.

(c) Consider the Binomial distribution in the case where $n \to \infty$, in such a manner that $nr = \lambda$. What happens to the probability generating function of the Binomial in this limit? Comment on what this has to say for the relationship between the Binomial distribution and the Poisson distribution.

> *Hint:* $\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$

(d) Use the probability generating function to show that the distribution of $N = N_1 + N_2$ is Poisson distributed with parameter $\lambda = \lambda_1 + \lambda_2$, where N_1 and N_2 are independent Poisson random variables with parameters λ_1 and λ_2, respectively.

Repeat the process using their distributions and convolution. Which way would you say is the preferred approach to finding the distribution of a sum of independent random variables of these two?

> *Hint:* the binomial theorem states that $\sum_{i=0}^{n} \frac{n!}{i!(n-i)!} a^i b^{(n-i)} = (a+b)^n$

Task 3: *Continuous time arrival process*

Assume boats are arriving at a region with the interval between them i.i.d. according to an exponential distribution with rate λ starting at time 0. That is, $p_{\Delta T_i}(\Delta t_i) = \lambda e^{-\lambda(\Delta t_i)}$, where ΔT_i is the time between the arrival of the ith and $i-1$th boat for $i > 0$. We let T_i denote the random arrival time of boat i so that $\Delta T_i = T_i - T_{i-1}$ for $i \geq 2$ and $\Delta T_1 = T_1$.

(a) At time $t_0 \geq 0$ we find that boat 1 has not yet arrived. Use Bayes rule to show that $p_{T_1 |\ T_1 \geq t_0}(t_1) = \lambda e^{-\lambda(t_1 - t_0)}$ for $t_1 \geq t_0$ and zero otherwise. Note that $p_{T_1}(t_1) = p_{\Delta T_1}(t_1) = \lambda e^{-\lambda(t_1)}$ from the problem setup.

> *Hint:* Remember that the Bayes rule is valid for all combinations of continuous and discrete random variables. You will need to find distributions over the binary event $T_1 \geq t_0$. Some might find it helpful to name this event and realization, say E and $e \in \{\text{True}, \text{False}\}$ respectively.

> *Note:* This result ($p_{X|X\geq a}(x) = p_X(x-a)$) is known as memorylessness, and is one of the properties that make the exponential distribution often used for modeling inter-arrival times in continuous time arrival processes. The exponential distribution is in fact the only continuous distribution to have this property. Can you think of why this property is called memorylessness, and why it is useful?

(b) Now with $T_0 = t_0$ for notational simplicity, what is the distribution of $T_n - T_0 = \sum_{i=1}^{n} T_i - T_{i-1} = \sum_{i=1}^{n} \Delta T_i$ given $T_1 \geq t_0$?

> *Hint:* Example 2.8

(c) What is the probability that boat $n+1$ did not arrive before $t > 0$ given that boat n arrived at time $T_n = t_n \leq t$? That is, find $\Pr(T_{n+1} > t | T_n = t_n \in [0, t])$.

> *Hint:*
>
> $$\begin{aligned} \Pr(T_{n+1} > t | T_n = t_n \in [0,t]) &= \Pr(T_{n+1} - t_n > t - t_n | T_n = t_n \in [0,t]) \\ &= \Pr(\Delta T_{n+1} > t - t_n | T_n = t_n \in [0,t]). \end{aligned}$$

Is there a dependence between ΔT_{n+1} and T_n (see task (a) and the problem statement)?

(d) Use the last two results to show that the distribution of the number of boats N to arrive between t_0 and t, given by $p_N(n) = \Pr(T_{n+1} > t \bigcap T_n \leq t | T_1 \geq t_0)$, is Poisson distributed.

> *Hint:* You should be able to form $p_{T_n, T_{n+1}}(t_n, t_{n+1} | T_1 \geq t_0)$ through conditioning and the results from (b) and (c), and then integrate t_n and t_{n+1} over a suitable region. The integration should not be very complicated after simplifying, and you should also be able to reuse some of the results in (c). Also, remember that $\Gamma(k) = (k-1)!$ for a positive integer k.

Task 4: *Finding posterior estimates of the number of boats in the region*

This is a conceptual continuation of the previous exercise.

A radar is installed to detect how many boats, n, there are in a region at a time $t > 0$. The radar scans the area, processes the data and reports how many boats it has counted, $m = n_D + m_{\text{fa}}$, where n_D is the correctly counted boats and m_{fa} is erroneous counts. Since the radar has been tuned to minimize the probability of counting something that is not a boat, it only detects and counts a boat with probability $P_{\text{D}} \in (0, 1)$ independently from all other counts.

m_{fa} is given to he Poisson distributed with parameter Λ and independent of all other counts.

Since each detection is Bernoulli distributed with parameter P_{D}, we have that $p(n_D|n) = \text{Binomial}(n_D; P_{\text{D}}, n) = \binom{n}{n_D} P_{\text{D}}^{n_D}(1 - P_{\text{D}})^{n-n_D}$. Letting $n_U = n - n_D$ we have in fact that $p(n_D, n_U|n) = p(n_D|n)$ since the binomial is a probabilistic unordered 'split' into the two categories detected and undetected.

n is given to be Poisson distributed with parameter λ.

Note: We have here used the implicit notation for PMFs and confused random variables with their realization (see the introduction to Section 2.2 and the end of sub-section 2.2.1). This means that $p(n_D|n)$ should be read as $p_{N_D|N}(n_D|n)$ where N is the random variable denoting the number of boats in the region and n is a realization and so on.

(a) Show that the marginal distribution for n_D and $n_U = n - n_D$ can be written as the product of two independent Poisson distributions $p(n_D, n_U) = p(n_D)p(n_U)$.

Hint: It might be easier to find $p(n_D, n)$ and then do a transform using $n = n_U + n_D$ afterwards. Since PMFs give probabilities, changing variables is simply done by algebraic substitution.

(b) Use the Bayes rule to show the number of detected boats is binomially distributed after you have received a measurement m, where $m = n_D + m_{\text{fa}}$. That is, find $p(n_D|m)$.

Hint: To find $p(m)$ you probably need to invoke an independence assumption and the result of task 2 (d).

(c) Find the MMSE and MAP estimate of n_D. You can use $\text{Binomial}(n_D; r, m)$ and insert for r afterward if you prefer.

Hint:
For MMSE: You might need $n\text{Binomial}(n; r, m) = mr\text{Binomial}(n - 1; r, m - 1)$ and the binomial theorem.

For MAP: Look at the sign of $p(n + 1) - p(n)$, and note how many peaks the distribution has. What can you say about this difference, specifically the sign, in relation to MAP and this/these peaks?