TTK4250 Sensor Fusion
Solution to Assignment 1

Task 1: *The CDF as a random variable*

Let X be a continuous-valued random variable with cumulative distribution function $P_X(\cdot) = \Pr(X \leq \cdot)$. 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: P_X in $P_X(X)$ should be interpreted as a function of real numbers or realizations. I.e. don't set the random variable X into $P_X(X)$ and evaluate the probability $\Pr(X \leq X)$, but consider $P_X(X)$ as a new stochastic variable. Keep in mind that $P_X(x)$ is a monotonous function, and that $P_X\big(P_X^{-1}(y)\big) = y$ 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
$\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} = \left(\frac{\mathrm{d}g(x)}{\mathrm{d}x}\big|_{x=g^{-1}(y)}\right)^{-1}$.

Solution: Let $P_Y(y)$ be the cdf of $Y = P_X(X)$. We then have

$$\begin{aligned} P_Y(y) &= \Pr(Y \leq y) = \Pr(P_X(X) \leq y) = \Pr\big(X \leq P_X^{-1}(y)\big) \\ &= P_X\big(P_X^{-1}(y)\big) = y, \end{aligned}$$

which is the cdf of a uniform random variable over $[0, 1]$, and concludes the expected solution.

Using the transformation of random variables formula becomes

$$p_Y(y) = p_X\big(P_X^{-1}(y)\big)\frac{\mathrm{d}}{\mathrm{d}y}P_X^{-1}(y) = p_X\big(P_X^{-1}(y)\big)\frac{1}{p_X\big(P_X^{-1}(y)\big)} = 1.$$

The second to last equality uses the calculus result stated above, and the fact that the derivative of a cdf is its pdf.

Great care should in general be taken when inverting functions like we did in the second equality. For a function to be invertible, it needs to be one-to-one. A cdf can have flat parts (pdf of zero) which would violate this. However, it is possible to define a (non-standard) appropriate inverse for this case. Can you think of how?

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 $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$.

Solution: From the definition of probability generating function, the distribution of a Poisson random variable and Taylor expansion of e^x we can calculate

$$GF_N(t) = E[t^N] = \sum_{n=0}^{\infty} t^n \frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} = e^{-\lambda} e^{t\lambda} = e^{\lambda(t-1)},$$

which is what we were supposed to show.

(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.

Solution: Again from the definition of generating functions, the distribution of a binomial random variable and the binomial theorem we have

$$GF_M(t) = E[t^M] = \sum_{m=0}^{n} \frac{n!}{m!(n-m)!} (tr)^m (1-r)^{n-m} = (1-r+rt)^n,$$

which is the desired result.

Using PGFs, the Binomial distribution is the sum of n i.i.d. Bernoulli distributions, that the PGF of a Bernoulli distribution is $GF_{\mathrm{Ber}(r)}(t) = 1-r+rt$ with property 2 of generating functions, we also get that $GF_M(t) = (GF_{\mathrm{Ber}(r)}(t))^n = (1-r+rt)^n$.

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

Solution: Inserting for r and using $\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$

$$\lim_{n\to\infty} GF_M(t) = \lim_{n\to\infty} (1-r+rt)^n = \lim_{n\to\infty} \left(1 + \frac{\lambda(t-1)}{n}\right)^n = e^{\lambda(t-1)}.$$

This is the generating function of a Poisson random variable, and we have thus shown that the binomial random variable is a Poisson random variable in the limit $n \to \infty$ such that $nr \to \lambda$. Note that this is in some sense a sum of infinitely many binary events all having 0 (or rather infinitesimal) probability of being true.

(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 $\sum_{i=0}^{n} \frac{n!}{i!(n-i)!} a^i b^{(n-i)} = (a+b)^n$

Solution: We first use PGFs. We know that when we have a sum of independent random variables, the sum's PGF becomes a product of the individual PGFs. Using this and knowing that $GF_{N_i}(t) = e^{\lambda_i(t-1)}$ we get the result

$$GF_N(t) = GF_{N_1}(t)GF_{N_2}(t) = e^{\lambda_1(t-1)}e^{\lambda_2(t-1)} = e^{(\lambda_1+\lambda_2)(t-1)} = e^{\lambda(t-1)},$$

Which is the GF of a Poisson random variable with parameter λ.

The convolution of two probability distributions gives the probability distribution of the sum of the two original random variables. If we try this we get

$$\begin{aligned}
p(N) &= p(N_1 + N_2) = p(N_1) * p(N_2) \\
&= \sum_{t=0}^{\infty} \frac{\lambda_1^t e^{-\lambda_1}}{t!} \frac{\lambda_2^{x-t} e^{-\lambda_2}}{(x-t)!} \\
&= e^{-(\lambda_1+\lambda_2)} \sum_{t=0}^{\infty} \frac{\lambda_1^t \lambda_2^{x-t}}{t!(x-t)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{x!} \sum_{t=0}^{x} \frac{x!}{t!(x-t)!} \lambda_1^t \lambda_2^{x-t} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{x!} (\lambda_1 + \lambda_2)^x
\end{aligned}$$

which is the Poisson distribution with parameter $\lambda_1 + \lambda_2$. Here we used the binominal series expansion:

$$(a+b)^x = \sum_{t=0}^{x} \frac{x!}{t!(x-t)!} a^t b^{x-t}$$

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?

Solution: First we have $\Pr(T_1 \geq t_0) = 1 - P_{T_1}(t_0)$ through the definition of cdf and probability of complementary sets, where $P_{T_1}(\cdot) = P_{\Delta T_1}(\cdot)$ is the cdf of T_1. $P_{T_1}(t_1) = 1 - e^{-\lambda t_1}$ is easily found through integration or looking it up. Secondly, $\Pr(T_1 \geq t_0 | T_1 = t_1)$ is deterministic; knowing T_1 we also know if it is larger or smaller than t_0 so the probabilities are simply 1 and 0. Further, we have

$$
\begin{aligned}
p_{T_1|T_1\geq t_0}(t_1) &= \frac{\Pr(T_1 \geq t_0 | T_1 = t_1) p_{T_1}(t_1)}{\Pr(T_1 \geq t_0)} \\
&= \begin{cases} \frac{p_{T_1}(t_1)}{1 - P_{T_1}(t_0)} = \frac{\lambda e^{-\lambda t_1}}{e^{-\lambda t_0}} = \lambda e^{-\lambda(t_1 - t_0)}, & t_1 \geq t_0 \\ \frac{0}{1 - P_{T_1}(t_0)} = 0, & t_1 < t_0 \end{cases}
\end{aligned}
$$

and shows the memorylessness property.

Since the conditional distribution has the exact same shape as the original, it 'forgets' how much time has gone after the event has been observed not to happen. That is, knowing how much time has passed does not influence the waiting time left for the event to happen. This creates independence between time intervals, which makes calculations easier. Of course, very few real-life processes truly possess this property, but it is in many cases a very good and convenient approximation.

(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

Solution: $T_n - T_0$ is a sum of n i.i.d. exponential random variables with parameter λ. From example 2.8 we know this to be a gamma distribution with shape parameter n and scale parameter $\frac{1}{\lambda}$. Otherwise, it can be shown by marginalization through

$$
\begin{aligned}
p_{T_n - T_0|T_1 > t_0}(t_n - t_0) &= \int_{t_0}^{t_n} \cdots \int_{t_0}^{t_3} \int_{t_0}^{t_2} \prod_{i=1}^{n} \lambda e^{-\lambda(t_i - t_{i-1})} \, \mathrm{d}t_1 \ldots \mathrm{d}t_{n-1} \\
&= \lambda^n e^{-\lambda(t_n - t_0)} \int_{t_0}^{t_n} \cdots \int_{t_0}^{t_3} \int_{t_0}^{t_2} \mathrm{d}t_1 \ldots \mathrm{d}t_{n-1} \\
&= \lambda^n e^{-\lambda(t_n - t_0)} \frac{(t_n - t_0)^{n-1}}{(n-1)!},
\end{aligned}
$$

which can be recognized as a gamma distribution with the given parameters.

(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)?

Solution: We get this probability using the hint, independence between ΔT_{n+1} and T_n, set complementary probability and the definition of the cdf

$$\begin{aligned}\Pr(T_{n+1} > t | T_n = t_n \in [0,t]) &= \Pr(\Delta T_{n+1} > t - t_n | T_n = t_n \in [0,t]) \\ &= \Pr(\Delta T_{n+1} > t - t_n) = 1 - \Pr(\Delta T_{n+1} \leq t - t_n) \\ &= 1 - P_{\Delta T_{n+1}}(t - t_n) = e^{-\lambda(t-t_n)}\end{aligned}$$

where $P_{\Delta T_{n+1}}(\cdot)$ is the cdf of ΔT_{n+1}.

(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.

Solution: Using conditioning, the hint can become

$$p_{T_n,T_{n+1}}(t_n, t_{n+1} | T_1 \geq t_0) = p_{T_{n+1}|T_n}(t_{n+1} | t_n, T_1 \geq t_0) p_{T_n}(t_n | T_1 \geq t_0).$$

Now $p_{T_{n+1}|T_n}(t_{n+1} | t_n, T_1 \geq t_0) = p_{T_{n+1}|T_n}(t_{n+1} | t_n)$ since T_1 has no information regarding T_{n+1} when T_n is given.

The region we want to integrate over is given by the task to be $T_n \in [0, t]$ and $T_{n+1} \in (t, \infty)$. But integrating $p_{T_{n+1}|T_n}$ over $T_{n+1} \in (t, \infty)$ gives the cdf and the answer from task (c).

Now using the result of task (b) gives

$$\begin{aligned}p_N(n) &= \int_{t_0}^{t} \Pr(t_{n+1} > t | T_n = t_n) p_{T_n|T_1 \geq t_0}(t_n) \, \mathrm{d}t_n \\ &= \int_{t_0}^{t} e^{-\lambda(t-t_n)} \lambda^n e^{-\lambda(t_n - t_0)} \frac{(t_n - t_0)^{n-1}}{(n-1)!} \, \mathrm{d}t_n \\ &= \int_{t_0}^{t} e^{-\lambda(t-t_0)} \lambda^n \frac{(t_n - t_0)^{n-1}}{(n-1)!} \, \mathrm{d}t_n \\ &= \frac{e^{-\lambda(t-t_0)} (\lambda(t - t_0))^n}{n!} = \text{Poisson}(n; \lambda(t - t_0)).\end{aligned}$$

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.

Solution: For $n < n_D$ the distribution is naturally 0, and for $n \geq n_D$ we have

$$\begin{aligned} p(n_D, n) &= \text{Binomial}(n_D; P_{\text{D}}, n)\text{Poisson}(n; \lambda) = \frac{n!}{n_D!(n - n_D)!} P_{\text{D}}^{n_D} (1 - P_{\text{D}})^{n-n_D} \frac{e^{-\lambda}\lambda^n}{n!} \\ &= \frac{e^{-\lambda P_{\text{D}}}(\lambda P_{\text{D}})^{n_D}}{n_D!} \frac{e^{-\lambda(1-P_{\text{D}})}(\lambda(1 - P_{\text{D}}))^{n-n_D}}{(n - n_D)!}. \end{aligned}$$

Letting $n_U = n - n_D$ we get what we are after,

$$p(n_D, n_U) = \text{Poisson}(n_D; \lambda P_{\text{D}})\text{Poisson}(n_U; \lambda(1 - P_{\text{D}})).$$

This tells us that the i.i.d. split of each object in a Poisson distributed count will result in two independent Poisson distributions. This is in many cases a simplifying property, such as in track initiation in some states of tracking algorithms.

(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).

Solution: We know that $n_D \sim \text{Poisson}(n_D; \lambda P_{\text{D}})$, and it is given that $m_{fa} \sim \text{Poisson}(m_{fa}; \Lambda)$. Independence and the result of task 2 (d) then gives that $m \sim \text{Poisson}(m; \lambda P_{\text{D}} + \Lambda)$. For $n_D \leq m$ we get, by using Bayes,

$$\begin{aligned} p(n_D|m) &= \frac{p(n_D, m)}{p(m)} = \frac{p(n_D)p(m|n_D)}{p(m)} = \frac{p(n_D)p(m_{fa})}{p(m)} \\ &= \frac{\frac{e^{-\lambda P_{\text{D}}}(\lambda P_{\text{D}})}{n_D!} \frac{e^{-\Lambda}\Lambda^{m-n_D}}{(m-n_D)!}}{\frac{e^{-(\lambda P_{\text{D}}+\Lambda)}(\lambda P_{\text{D}}+\Lambda)^m}{m!}} \\ &= \frac{m!}{n_D!(m - n_D)!} \left(\frac{\lambda P_{\text{D}}}{\lambda P_{\text{D}} + \Lambda}\right)^{n_D} \left(\frac{\Lambda}{\lambda P_{\text{D}} + \Lambda}\right)^{m-n_D}, \end{aligned}$$

which is a Binomial distribution with probability parameter $\frac{\lambda P_\mathrm{D}}{\lambda P_\mathrm{D}+\Lambda}$, and number parameter m. This can naturally be read as each count being Bernoulli distributed for being a boat or false alarm.

This is exactly the distribution — although sometimes generalized to be dependent on the "measurement position" — used to model initial track existence in some state-of-the-art target tracking algorithms.

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

Hint:
For MMSE: You might need $n\mathrm{Binomial}(n; r, m) = mr\mathrm{Binomial}(n-1; r, m-1)$ 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?

Solution: From theorem 2.6.1 we know that the MMSE is given by the expected value of the posterior distribution. We thus calculate the expected value of the Binomial distribution we found above

$$\begin{aligned}
E[n_D] &= \sum_{n_D=0}^{m} n_D \frac{m!}{n_D!(m-n_D)!} r^{n_D}(1-r)^{m-n_D} \\
&= rm \sum_{n_D=1}^{m} \frac{(m-1)!}{(n_D-1)!(m-n_D)!} r^{n_D-1}(1-r)^{m-n_D} \\
&= rm \sum_{i=0}^{m-1} \frac{(m-1)!}{i!(m-1-i)!} r^{i}(1-r)^{m-1-i} = rm = m\frac{\lambda P_\mathrm{D}}{\lambda P_\mathrm{D}+\Lambda}
\end{aligned}$$

For the MAP we can simply find the peak, or mode if you like, of the posterior distribution. We do this by noting that the distribution has a single peak and finding the first n_D for which the difference $p(n_D+1) - p(n_D)$ is negative. To simplify notation we use n and r as parameters:

$$\begin{aligned}
p(n+1) - p(n) &= \frac{m!}{(n+1)!(m-(n+1))!} r^{n+1}(1-r)^{m-(n+1)} - \frac{m!}{n!(m-n)!} r^{n}(1-r)^{m-n} \\
&= \frac{m!}{(n+1)!(m-n)!} r^{n}(1-r)^{m-(n+1)} \left[(m-n)r - (n+1)(1-r)\right] \\
&= \frac{m!}{(n+1)!(m-n)!} r^{n}(1-r)^{m-(n+1)} \left[-n + (m+1)r - 1\right]
\end{aligned}$$

We see that the bracketed term is positive for $n < (m+1)r - 1$, zero for $n = (m+1)r - 1$ and negative for $n > (m+1)r - 1$. Since n is discrete we have that $n = \lceil (m+1)r - 1 \rceil = \lfloor (m+1)r \rfloor = \lfloor (m+1)\frac{\lambda P_\mathrm{D}}{\lambda P_\mathrm{D}+\Lambda} \rfloor$ is our MAP estimate. However, if $(m+1)r$ is integer, we have that $n = (m+1)r - 1$ has the same pmf value and is therefore also a valid MAP estimate.