Problem 1

For $t \in R^{\geq 0}$ be X_t a random variable having normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ (this is written as $X_t \sim N(0,1)$) and furthermore all X_t are independent. We define the stochastic process $(Y_t)_{t \in R^{\geq 0}}$ by: $Y_t = t + X_t$.

a. Is $(Y_t)_{t \in R^{\geq 0}}$ stationary ?

b. Has $(Y_t)_{t \in R^{\geq 0}}$ independent increments ?

c. Has $(Y_t)_{t \in R^{\geq 0}}$ stationary increments ?

d. Find $E[Y_t]$ and $Var[Y_t]$.

The definition of the term "stationary" requires that the distribution fulfills the condition:

$$F_{Y(t_1),\ldots,Y(t_n)}(x_1,\ldots,x_n) = F_{Y(t_1+s),\ldots,Y(t_n+s)}(x_1,\ldots,x_n)$$

for all $n \in N, s \in R$. For $n = 1$ the expression becomes even more handy:

$$F_{Y(t_1)}(x) = F_{Y(t_1+s)}(x)$$

If $(Y_t)_{t \in R^{\geq 0}}$ is stationary then it should be true that:

$$\begin{aligned} \Pr[Y_t \leq x] &= \Pr[Y_{t+s} \leq x] \\ \Pr[X_t + t \leq x] &= \Pr[X_{t+s} + t + s \leq x] \\ \Pr[N(0,1) \leq x - t] &= \Pr[N(0,1) \leq x - t - s] \end{aligned}$$

Unfortunately, the last line offers a contradiction. Therefore we proved that $(Y_t)_{t \in R^{\geq 0}}$ is not stationary.

For independent increments the random variables $X(t_1) - X(t_0)$, $X(t_2) - X(t_1)$, ... $X(t_n) - X(t_{n-1})$ are independent for all $n \in N$ and $t_i \in T$ (under the assumption that all t_i are in ascending order which means $t_0 < t_1 < \ldots < t_{n-1} < t_n$).

The first step is:

$$\begin{aligned} Y_{t_0+s} - Y_{t_0} &= X_{t_0+s} + t_0 + s - (X_{t_0} + t_0) \\ &= X_{t_0+s} + s - X_{t_0} \end{aligned}$$

For any subsequent increment:

$$\begin{aligned} Y_{t_0+s+r} - Y_{t_0+r} &= X_{t_0+s+r} + t_0 + s + r - (X_{t_0} + t_0 + r) \\ &= X_{t_0+s} + s - X_{t_0} \end{aligned}$$

It was stated that all X_t are independent. Therefore, all increments $X_0 - X_1$, $X_1 - X_2$, etc. have to be independent, too. Because s is a constant value it does not influence any dependency and thus $Y_0 - Y_1$, $Y_1 - Y_2$, etc. are independent as well. That result leads to the conclusion that $(Y_t)_{t \in R^{\geq 0}}$ has independent increments.

Stationary increments need to accomplish:

$$Y(t_2 + s) - Y(t_1 + s) \stackrel{st}{=} Y(t_2) - Y(t_1)$$

Replacing Y_t by X_t yields:

$$\begin{aligned}
& Y(t_2 + s) - Y(t_1 + s) \\
& \stackrel{st}{=} N(t_2 + s, 1) - N(t_1 + s, 1) \\
& \stackrel{st}{=} N\left(t_2 - t_1, \sqrt{2}\right) \\
& \stackrel{st}{=} N(t_2, 1) - N(t_1, 1) \\
& \stackrel{st}{=} Y(t_2) - Y(t_1)
\end{aligned}$$

Hereby we have shown $(Y_t)_{t \in R^{\geq 0}}$ produces stationary increments.

The expectation value is given by $E[Y_t] = E[X_t + t] = EX_t + t = 0 + t = t$ (remember: $X_t \sim N(0,1)$)
Furthermore $Var[Y_t] = 1$.

Problem 2

Be $\left(N(t)\right)_{t\in R^{\geq 0}}$ a Poisson Process with rate λ. We are given that one arrival occurred in the interval $(0,t]$ (hence $N(t)=1$) but we do not know when this arrival happened. Let Y denote the random variable of the first arrival time, the range of Y is $(0,t]$. Show that Y has a uniform distribution (hence, its distribution function is given by $F(y)=\Pr[Y\leq y]=\frac{y}{t}$). Hint: compute $\Pr[Y\leq y|N(t)=1]$.

An interesting property of the Poisson process is:

$$\Pr[N(t)=k]=\frac{(\lambda\cdot t)^k e^{-\lambda t}}{k!}$$

The law of conditional probabilities gives:

$$\begin{aligned}\Pr[Y\leq y|N(t)=1] &= \frac{\Pr[Y\leq y, N(t)=1]}{\Pr[N(t)=1]}\\ &= \frac{\Pr[N(t)=1]\cdot\Pr[N(t-y)=0]}{\Pr[N(t)=1]}\end{aligned}$$

Applying the formula developed above gives:

$$\begin{aligned}&= \frac{e^{-\lambda y}\cdot\lambda y\cdot e^{-\lambda\cdot(t-y)}}{e^{-\lambda t}\cdot\lambda t}\\ &= \frac{e^{-\lambda y}\cdot\lambda y\cdot e^{-\lambda t}\cdot e^{\lambda y}}{e^{-\lambda t}\cdot\lambda t}\\ &= \frac{y}{t}\end{aligned}$$

In conclusion, we showed that if $N(t)=1$ occurs in the interval $(0,t]$ then Y is uniformly distributed.

Problem 3

Be $(N(t))_{t\in R^{\geq 0}}$ a Poisson Process with rate λ, and Y a continuous non-negative random variable independent of $(N(t))_{t\in R^{\geq 0}}$ with density function $f(y)$.

a. *Show that $G_{N(Y)}(z) = E\left[z^{N(Y)}\right]$ can be expressed as $G_{N(Y)}(z) = M_Y(-\lambda\cdot(1-z))$, i.e. the z-Transform of $N(Y)$ can be expressed in terms of the moment generating function for Y. Hint: law of total probability for continuous random variables.*

b. *Use $G_{N(Y)}(z)$ to find $E[N(Y)]$ and $Var[N(Y)]$.*

Obviously:

$$G_{N(Y)}(z) = E\left[z^{N(Y)}\right]$$
$$= \sum_{n=0}^{\infty} z^n \cdot \Pr[N(Y)=n]$$

That can be expressed as:

$$\sum_{n=0}^{\infty} z^n \cdot \Pr[N(Y)=n] = \sum_{n=0}^{\infty} z^n \cdot \int_0^{\infty} \Pr[N(Y)=n | Y=\tau]\cdot f(\tau)d\tau$$
$$= \sum_{n=0}^{\infty} z^n \cdot \int_0^{\infty} \Pr[N(\tau)=n]\cdot f(\tau)d\tau$$
$$= \sum_{n=0}^{\infty} z^n \cdot \int_0^{\infty} e^{-\lambda\tau} \cdot \frac{(\lambda\tau)^n}{n!}\cdot f(\tau)d\tau$$
$$= \sum_{n=0}^{\infty} \int_0^{\infty} z^n \cdot e^{-\lambda\tau} \cdot \frac{(\lambda\tau)^n}{n!}\cdot f(\tau)d\tau$$

Because of $\int_a^b \sum_{n=0}^{\infty} f_n(x)dx = \sum_{n=0}^{\infty} \int_a^b f_n(x)dx$

$$\sum_{n=0}^{\infty} \int_0^{\infty} z^n \cdot e^{-\lambda\tau} \cdot \frac{(\lambda\tau)^n}{n!}\cdot f(\tau)d\tau = \int_0^{\infty} \sum_{n=0}^{\infty} e^{-\lambda\tau} \cdot z^n \cdot \frac{(\lambda\tau)^n}{n!}\cdot f(\tau)d\tau$$
$$= \int_0^{\infty} f(\tau)\cdot e^{-\lambda\tau} \cdot \sum_{n=0}^{\infty} \frac{(z\lambda\tau)^n}{n!}\cdot d\tau$$

According to Taylor's observations the following holds: $\sum_{n=0}^{\infty}\frac{x^n}{n!}=e^x$. Hence, we get:

$$\begin{aligned}\int_0^{\infty} f(\tau)\cdot e^{-\lambda\tau}\cdot\sum_{n=0}^{\infty}\frac{(z\lambda\tau)^n}{n!}\cdot d\tau &= \int_0^{\infty} f(\tau)\cdot e^{-\lambda\tau}\cdot e^{z\lambda\tau}d\tau \\ &= \int_0^{\infty} f(\tau)\cdot e^{\lambda\tau(-1+z)}d\tau \\ &= \int_0^{\infty} f(\tau)\cdot e^{-\lambda\cdot(1-z)\tau}d\tau \\ &= M_Y(-\lambda\cdot(1-z))\end{aligned}$$

The computation of the expectation value $E[N(Y)]=G'_{N(Y)}(1)$ yields:

$$G'_{N(Y)}(z)=\left(\sum_{n=0}^{\infty}z^n\cdot\int_0^{\infty}e^{-\lambda\tau}\cdot\frac{(\lambda\tau)^n}{n!}\cdot f(\tau)d\tau\right)'$$

The first derivation of a sum leads to:

$$\left(\sum_{n=0}^{\infty}z^n\right)'=\sum_{n=1}^{\infty}n\cdot z^{n-1}$$

Using that formula gives:

$$\begin{aligned}\left(\sum_{n=0}^{\infty}z^n\cdot\int_0^{\infty}e^{-\lambda\tau}\cdot\frac{(\lambda\tau)^n}{n!}\cdot f(\tau)d\tau\right)' &= \sum_{n=1}^{\infty}n\cdot z^{n-1}\cdot\int_0^{\infty}e^{-\lambda\tau}\cdot\frac{(\lambda\tau)^n}{n!}\cdot f(\tau)d\tau \\ &= \int_0^{\infty}e^{-\lambda\tau}\cdot f(\tau)\cdot\sum_{n=1}^{\infty}\left(n\cdot\frac{(\lambda\tau)^n}{n!}\cdot z^{n-1}\right)d\tau \\ &= \int_0^{\infty}e^{-\lambda\tau}\cdot f(\tau)\cdot\lambda\tau\cdot\sum_{n-1=0}^{\infty}\left(\frac{(\lambda\tau z)^{n-1}}{(n-1)!}\right)d\tau \\ &= \int_0^{\infty}e^{-\lambda\tau}\cdot f(\tau)\cdot\lambda\tau\cdot e^{\lambda\tau z}\,d\tau\end{aligned}$$

That integral can be used to determine the expectation value:

$$\begin{aligned}E[N(Y)] &= G'_{N(Y)}(1) \\ &= \int_0^{\infty}e^{-\lambda\tau}\cdot f(\tau)\cdot\lambda\tau\cdot e^{\lambda\tau\cdot 1}\,d\tau \\ &= \int_0^{\infty}\lambda\tau\cdot f(\tau)\,d\tau=\lambda\int_0^{\infty}\tau\cdot f(\tau)\,d\tau \\ &= \lambda E[Y]\end{aligned}$$

The next step, calculating $Var[N(Y)]$, involves the utilization of:

$$Var[N(Y)] = G''_{N(Y)}(1) + G'_{N(Y)}(1) - \left(G'_{N(Y)}(1)\right)^2$$

Again, some derivations have to be done:

$$\left(\sum_{n=0}^{\infty} z^n\right)'' = \left(\sum_{n=1}^{\infty} n \cdot z^{n-1}\right)'$$
$$= \sum_{n=2}^{\infty} n \cdot (n-1) \cdot z^{n-2}$$

We find:

$$G''_{N(Y)}(z) = \left(\sum_{n=0}^{\infty} z^n \cdot \int_0^{\infty} e^{-\lambda\tau} \cdot \frac{(\lambda\tau)^n}{n!} \cdot f(\tau)\,d\tau\right)''$$
$$= \sum_{n=2}^{\infty} n \cdot (n-1) \cdot z^{n-2} \int_0^{\infty} e^{-\lambda\tau} \cdot \frac{(\lambda\tau)^n}{n!} \cdot f(\tau)\,d\tau$$
$$= \int_0^{\infty} e^{-\lambda\tau} \cdot (\lambda\tau)^2 \cdot f(\tau) \cdot \sum_{n-2=0}^{\infty} \frac{(\lambda\tau z)^{n-2}}{(n-2)!}\,d\tau$$
$$= \int_0^{\infty} e^{-\lambda\tau} \cdot (\lambda\tau)^2 \cdot f(\tau) \cdot e^{\lambda\tau z}\,d\tau$$

And finally:

$$Var[N(Y)] = G''_{N(Y)}(1) + G'_{N(Y)}(1) - \left(G'_{N(Y)}(1)\right)^2$$
$$= \int_0^{\infty} (\lambda\tau)^2 \cdot f(\tau)\,d\tau + \int_0^{\infty} \lambda\tau \cdot f(\tau)\,d\tau - \left(\int_0^{\infty} \lambda\tau \cdot f(\tau)\,d\tau\right)^2$$
$$= \lambda^2 \int_0^{\infty} \tau^2 \cdot f(\tau)\,d\tau + \lambda \int_0^{\infty} \tau \cdot f(\tau)\,d\tau - \left(\lambda \int_0^{\infty} \tau \cdot f(\tau)\,d\tau\right)^2$$
$$= \lambda^2 E[Y^2] + \lambda E[Y] - (\lambda E[Y])^2$$
$$= \lambda E[Y] + \lambda^2 \left(E[Y^2] - (E[Y])^2\right)$$
$$= \lambda E[Y] + \lambda^2 Var[Y]$$

Problem 4

Sometimes one is interested in making predictions, e.g. we know that a file transfer already took t time units, what is the probability that it takes another $d > 0$ time units ? A random variable X is called heavy-tailed, when $\Pr[X > t] \sim t^{-\alpha}$ $(t \to \infty)$ for $0 < \alpha < 2$. A function $f(x)$ behaves asymptotically as $g(x)$, written as $f(x) \sim g(x)$, if there exists some $c \in R, c \neq 0$ such that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = c$ holds, $g(x) \neq 0$ required.

a. *Show that the exponential distribution with parameter λ is not heavy-tailed.*

b. *Compute $\lim_{t\to\infty} \Pr[X > t + d | X > t]$ for the exponential distribution.*

c. *Compute $\lim_{t\to\infty} \Pr[X > t + d | X > t]$ for the heavy-tailed distribution.*

d. *Compare and interpret the results*

In many studies it has been found that the distribution of file sizes in a UNIX file system and the document sizes of a web server are heavy-tailed.

The exponential distribution is determined by:

$$\begin{aligned}\Pr[Exp(\lambda) \leq t] &= \int_0^t \lambda \cdot e^{-\lambda x}\, dx \\ &= 1 - e^{-\lambda x}\end{aligned}$$

Because of $\Pr[X \leq x] = 1 - \Pr[X > x]$:

$$\begin{aligned}\Pr[Exp(\lambda) > t] &= 1 - \left(1 - e^{-\lambda x}\right) \\ &= e^{-\lambda x}\end{aligned}$$

If $f(x) \sim g(x)$ we have to find some $\lim_{x\to\infty} \frac{f(x)}{g(x)} = c$ where $f(x) = e^{-\lambda x}$ and $g(x) = x^{-\alpha}$:

$$\begin{aligned}\lim_{t\to\infty} \frac{e^{-\lambda t}}{t^{-\alpha}} &= \lim_{t\to\infty} \frac{t^{\alpha}}{e^{\lambda t}} \\ &= \lim_{t\to\infty} \frac{t^{\alpha}}{\left(e^{\lambda}\right)^{t}} \\ &= 0\end{aligned}$$

since α is fixed. As a result, there does not exist some $c \neq 0$ and hence the exponential distributed is not heavy-tailed, too.

A very handy property of the exponential distribution, its memorylessness, allows us to determine the probability in just three lines:

$$\begin{aligned}\lim_{t\to\infty}\Pr\left[Exp(\lambda) > t + d \middle| Exp(\lambda) > t\right] &= \lim_{t\to\infty}\Pr\left[Exp(\lambda) > d\right]\\ &= \Pr\left[Exp(\lambda) > d\right]\\ &= e^{-\lambda d}\end{aligned}$$

For the heavy-tailed distribution:

$$\begin{aligned}\lim_{t\to\infty}\Pr\left[X > t + d \middle| X > t\right] &= \lim_{t\to\infty}\frac{\Pr\left[X > t + d, X > t\right]}{\Pr\left[X > t\right]}\\ &= \lim_{t\to\infty}\frac{\Pr\left[X > t + d\right]}{\Pr\left[X > t\right]}\\ &= \lim_{t\to\infty}\frac{(t+d)^{-\alpha}}{t^{-\alpha}}\\ &= \lim_{t\to\infty}\left(\frac{t}{t+d}\right)^{\alpha}\\ &= 1^{\alpha}\\ &= 1\end{aligned}$$

We did not have the time to translate our conclusions – there are still written in German:

Für die Exponentialverteilung nimmt die Wahrscheinlichkeit dafür, dass ein Übertragungsvorgang noch eine Zeitspanne d anhält, unabhängig vom betrachteten Zeitpunkt, während der Übertragung exponentiell mit der Länge der Zeitspanne ab. Damit wird eine kurze Zeitspanne d favorisiert, unabhängig von der schon verstrichenen Zeit. Das bedeutet, dass das Ende der Übertragung immer höchstwahrscheinlich nahe ist.

Für eine heavy-tailed-Verteilung ist die Wahrscheinlichkeit dafür, dass ein Übertragungsvorgang anhält, immer nahe Eins, unabhängig vom betrachteten Zeitpunkt und der Länge des Zeitfensters. Somit wird vorhergesagt, dass die Übertragung immer weiter andauert und das Ende der Übertragung eher unwahrscheinlich ist.

→ Kurze Dateiübertragungen lassen sich recht gut mit der Exponentialverteilung beschreiben, lange Dateiübertragungen entsprechen dagegen eher einer heavy-tailed-Verteilung.

Problem 5 (bonus)

The distribution and density of a random variable X with Pareto distribution are given by:

$$F(x) = \Pr[X \le x] = 1 - \left(\frac{k}{x}\right)^{\alpha}$$

$$f(x) = \alpha \cdot k^{\alpha} \cdot x^{-\alpha-1}$$

for $\alpha > 0$, $k > 0$ and $x \ge k$. For $0 < \alpha < 2$ the Pareto distribution is heavy-tailed. Be X such a Pareto random variable with parameters $\alpha = 1$, $k = 1$. Furthermore, be Y an exponential random variable with parameter $\lambda = 1$. Plot both $\Pr[X > x]$ and $\Pr[Y > x]$ for $1 \le x \le 100$ using a doubly-logarithmic scale.