Seminar Nr. 5, Continuous Random Variables and Continuous Random Vectors

Theory Review

$X : S \to \mathbb{R}$ continuous random variable with pdf $f : \mathbb{R} \to \mathbb{R}$ and cdf $F : \mathbb{R} \to \mathbb{R}$. Properties:

1. $F(x) = P(X \le x) = \int\limits_{-\infty}^{x} f(t)dt$

2. $f(x) \ge 0, \forall x \in \mathbb{R}, \int\limits_{\mathbb{R}} f(x) = 1$

3. $P(X = x) = 0, \forall x \in \mathbb{R}, P(a < X < b) = P(a \le X \le b) = \int\limits_{a}^{b} f(t)dt$

4. $F(-\infty) = 0, F(\infty) = 1$

$(X, Y) : S \to \mathbb{R}^2$ continuous random vector with pdf $f = f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$ and

cdf $F = F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R},\ F(x, y) = P(X \le x, Y \le y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(u, v)\ dv\ du,\ \forall (x, y) \in \mathbb{R}^2$. Properties:

1. $P\left(a_1 < X \le b_1, a_2 < Y \le b_2\right) = F\left(b_1, b_2\right) - F\left(a_1, b_2\right) - F\left(b_1, a_2\right) + F\left(a_1, a_2\right)$
2. $F(\infty, \infty) = 1, F(-\infty, y) = F(x, -\infty) = 0,\ \forall x, y \in \mathbb{R}$
3. $F_X(x) = F(x, \infty),\ F_Y(y) = F(\infty, y),\ \forall x, y \in \mathbb{R}$ (marginal cdf's)
4. $P((X, Y) \in D) = \iint\limits_{D} f(x, y)\ dy\ dx$
5. $f_X(x) = \int\limits_{\mathbb{R}} f(x, y)dy,\ \forall x \in \mathbb{R}, f_Y(y) = \int\limits_{\mathbb{R}} f(x, y)dx,\ \forall y \in \mathbb{R}$ (marginal densities)
6. X and Y are independent <=> $f_{(X,Y)}(x, y) = f_X(x)f_Y(y),\ \forall (x, y) \in \mathbb{R}^2$.

Function $Y = g(X)$: X r.v., $g : \mathbb{R} \to \mathbb{R}$ differentiable with $g' \ne 0$, strictly monotone

$f_Y(y) = \dfrac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|},\ y \in g\left(\mathbb{R}\right)$

Uniform distribution $U(a, b),\ -\infty < a < b < \infty$: pdf $f(x) = \dfrac{1}{b - a}, x \in [a, b]$.

Normal distribution $N(\mu, \sigma), \mu \in \mathbb{R}, \sigma > 0$: pdf $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$.

Gamma distribution $Gamma(a, b),\ a, b > 0$: pdf $f(x) = \dfrac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}}\ ,\ x > 0$.

Exponential distribution $Exp(\lambda) = Gamma(1, 1/\lambda),\ \lambda > 0$: pdf $f(x) = \lambda e^{-\lambda x}, x > 0$.
- Exponential distribution models *time*: waiting time, interarrival time, failure time, time between rare events, etc; the parameter λ represents the frequency of rare events, measured in time^{-1}.
- Gamma distribution models the *total* time of a multistage scheme.
- For $\alpha \in \mathbb{N}$, a $Gamma(\alpha, 1/\lambda)$ variable is the sum of α independent $Exp(\lambda)$ variables.

---

1. The lifetime, in years, of some electronic component is a random variable with density

$$f(x) = \begin{cases} \dfrac{k}{x^4}, & \text{for} \quad x \ge 1 \\ 0, & \text{for} \quad x < 1. \end{cases}$$

Find
a) the constant k;
b) the corresponding cdf F;
c) the probability for the lifetime of the component to exceed 2 years.

Solution:

a) We find the constant k from the condition $\int_{\mathbb{R}} f(x)\,dx = 1$. We have

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= k\int_{1}^{\infty} \frac{1}{x^4}dx = k\int_{1}^{\infty} x^{-4}\,dx \\ &= -\frac{k}{3}\,\frac{1}{x^3}\Big|_{1}^{\infty} = \frac{k}{3} = 1,\end{aligned}$$

so $k = 3$ and

$$f(x) = \frac{3}{x^4},\ x \geq 1.$$

b) The cdf is found by

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

If $x < 1$, we integrate the constant 0, so $F(x) = 0$. For $x \geq 1$, we have

$$\begin{aligned}F(x) &= \int_{-\infty}^{x} f(t)dt = \int_{1}^{x} \frac{3}{t^4}dt \\ &= 3\int_{1}^{x} t^{-4}\,dt = -\frac{1}{t^3}\Big|_{1}^{x} = 1 - \frac{1}{x^3}.\end{aligned}$$

Thus,

$$F(x) = \begin{cases} 0, & x < 1 \\ 1 - \dfrac{1}{x^3}, & x \geq 1. \end{cases}$$

c) Recall that X is the *lifetime*, in years, of the component. So, we want

$$\begin{aligned}P(X > 2) &= 1 - P(X \leq 2) = 1 - F(2) \\ &= 1 - \left(1 - \frac{1}{8}\right) = \frac{1}{8} = 0.125.\end{aligned}$$

2. (The Uniform property) Let $X \in U(a,b)$. For any $h > 0$ and $t, s \in [a, b-h]$,

$$P(s < X < s+h) = P(t < X < t+h).$$

The probability is only determined by the length of the interval, but not by its location.
Example: A certain flight can arrive at any time between 4:50 and 5:10 pm. Let X denote the arrival time of the flight.
a) What distribution does X have?

b) When is the flight more likely to arrive: between 4:50 and 4:55 or between 5 and 5:05; before 4:55 or after 5:05?

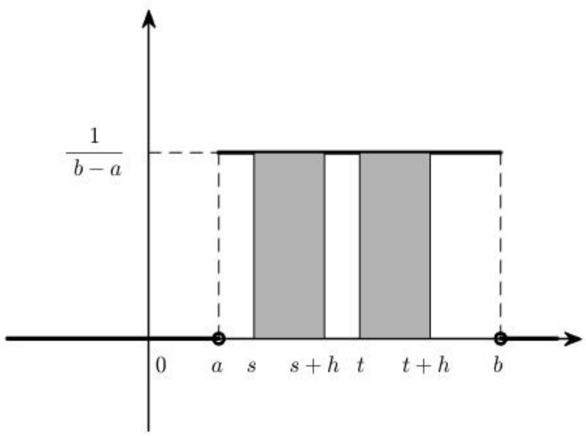
Solution:
Recall that

$$P(a < X \le b) \;=\; F(b) - F(a) \;=\; \int_a^b f(t)dt$$

and the interval can be open or closed at either endpoint. Then

$$\begin{aligned} P(s < X < s+h) &= \int_s^{s+h} \frac{1}{b-a}dx &= \frac{1}{b-a}x\Big|_s^{s+h} &= \frac{h}{b-a} \\ P(t < X < t+h) &= \int_t^{t+h} \frac{1}{b-a}dx &= \frac{1}{b-a}x\Big|_t^{t+h} &= \frac{h}{b-a}. \end{aligned}$$

Also, we can see it graphically. Each probability is the area of a rectangle whose height is $\frac{1}{b-a}$. So, as long as the two rectangles have also the same *base*, the areas will be equal, no matter where the starting point s (or t) is in the interval $[a, b]$.



Example:
a) Since the flight can arrive at *any* time in that interval, it means the time is Uniformly distributed in the interval $[4:50, 5:10]$, so $X \in U(4:50, 5:10)$.
b) The flight is *just as likely* to arrive between 4:50 and 4:55 as it is to arrive between 5 and 5:05, because the two intervals have the *same length*.
Same for before 4:55 or after 5:05, just as likely, for the same reason.

3. On the average, a computer experiences breakdowns every 5 months. The time until the first breakdown and the times between any two consecutive breakdowns are independent Exponential random variables. After the third breakdown, a computer requires a special maintenance.
a) Find the probability that a special maintenance is required within the next 9 months;
b) Given that a special maintenance was not required during the first 12 months, what is the probability that it will not be required within the next 4 months?

Solution:
Since computer breakdowns qualify as rare events (they cannot occur simultaneously), the time between two consecutive breakdowns has Exponential distribution. Since breakdowns occur every 5 months, their frequency is $\lambda = 1/5$ per month. So the distribution is $Exp(1/5)$. Now, since breakdowns are independent of each other, the time T until the third breakdown (when a special maintenance is required) is the sum of 3 $Exp(1/5)$ variables and,

hence, has a $Gamma(3,5)$ distribution.
a) Then we want

$$P(T \le 9) \;=\; F_T(9) \;\overset{\text{Matlab}}{=}\; gamcdf(9,3,5) \;=\; 0.2694.$$

b) This is *conditional* probability.

$$\begin{aligned} P(T > 12 + 4 \,|\, T > 12) &= \frac{P(T > 16)}{P(T > 12)} \;=\; \frac{1 - P(T \le 16)}{1 - P(T \le 12)} \\ &= \frac{1 - F_T(16)}{1 - F_T(12)} \;\overset{\text{Matlab}}{=}\; \frac{1 - gamcdf(16,3,5)}{1 - gamcdf(12,3,5)} \;=\; 0.6668. \end{aligned}$$

4. The joint density for (X, Y) is $f_{(X,Y)}(x, y) = \dfrac{1}{16}x^3y^3$, $x, y \in [0, 2]$.
a) Find the marginal densities f_X, f_Y.
b) Are X and Y independent?
c) Find $P(X \le 1)$.

Solution:

We find the marginal pdf of one of the component by integrating over $\mathbb{R}$ with respect to the *other* variable. Let us consider f_X first. In order for the vector to have a nonzero pdf, namely $\dfrac{1}{16}x^3y^3$, *both* x and y have to be in $[0, 2]$. So, if $x \notin [0, 2]$, we integrate the function 0 and, thus, $f_X(x) = 0$. For $x \in [0, 2]$, we have

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f(x, y)\, dy \;=\; \frac{1}{16}\int_0^2 x^3y^3 dy \\ &= \frac{x^3}{16}\int_0^2 y^3\, dy \;=\; \frac{x^3}{16}\cdot\frac{1}{4}y^4\Big|_{y=0}^{y=2} \;=\; \frac{1}{4}x^3. \end{aligned}$$

Thus,

$$f_X(x) \;=\; \frac{1}{4}x^3,\ x \in [0, 2].$$

By symmetry (both the function $f(x, y)$ and the domain $[0, 2] \times [0, 2]$ are symmetric), we get the same result for f_Y. So

$$f_Y(y) \;=\; \frac{1}{4}y^3,\ y \in [0, 2].$$

b) We check that

$$f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y), \forall (x, y) \in \mathbb{R}^2.$$

First, the trivial case. If $x \notin [0, 2]$ or $y \notin [0, 2]$ (or both), the left-hand side is 0 and at least one of the factors on the right is 0, so we have equality $0 = 0$.
If both $x, y \in [0, 2]$, then we have

$$\frac{1}{16}x^3y^3 \;=\; \frac{1}{4}x^3 \cdot \frac{1}{4}y^3,$$

equality again, so, yes, they are independent.

c)

$$\begin{aligned} P(X \le 1) &= \int_{-\infty}^{1}\int_{-\infty}^{\infty} f(x, y)\, dydx \;=\; \int_{-\infty}^{1} f_X(x)\, dx \\ &= \frac{1}{4}\int_0^1 x^3\, dx \;=\; \frac{1}{4}\cdot\frac{1}{4}x^4\Big|_0^1 \;=\; \frac{1}{16}. \end{aligned}$$

5. Let X be a random variable with density $f_X(x) = \dfrac{1}{4} x e^{-\frac{x}{2}}$, $x \geq 0$ and let $Y = \dfrac{1}{2}X + 2$. Find f_Y.

Solution:

Here, $Y = g(X)$, with $g : \mathbb{R} \to \mathbb{R}$

$$g(x) = \frac{1}{2}x + 2.$$

So, first off, since g is a linear function, its range is $g(\mathbb{R}) = \mathbb{R}$.

Next, we check the conditions of the theorem on the function g. We have $g'(x) = \dfrac{1}{2}$, which exists everywhere and is never equal to 0. Moreover, since $g'(x) = \dfrac{1}{2} > 0$, $\forall x \in \mathbb{R}$, g is *strictly increasing*. Thus, it satisfies all the hypotheses and

$$f_Y(y) = \frac{f_X\left(g^{-1}(y)\right)}{|g'\left(g^{-1}(y)\right)|},$$

for every $y \in \mathbb{R}$.

Since the derivative of g is a constant, the denominator is simply equal to $\dfrac{1}{2}$.

For the numerator, we need the *inverse* of g. Recall we find the inverse by solving for x

$$y = g(x) \iff x = g^{-1}(y).$$

We have

$$\begin{aligned} \frac{1}{2}x + 2 &= y \\ \frac{1}{2}x &= y - 2 \\ x &= 2(y-2). \end{aligned}$$

So

$$g^{-1}(y) = 2(y-2).$$

Then the numerator will be (simple composition of functions)

$$\begin{aligned} f_X\left(g^{-1}(y)\right) &= f_X\left(2(y-2)\right), \\ &= \frac{1}{4}\big(2(y-2)\big)e^{-\frac{(2(y-2))}{2}},\ 2(y-2) \geq 0, \\ &= \frac{1}{2}(y-2)e^{-(y-2)}, y \geq 2. \end{aligned}$$

Thus,

$$f_Y(y) = (y-2)e^{2-y}, y \geq 2 \ \text{(and } 0, \text{ otherwise).}$$

6. Let $X \in N(0,1)$. Find the probability density function of $Y = |X|$.

Solution:

For $X \in N(0,1)$, the pdf is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},\ \forall x \in \mathbb{R}.$$

Main idea: Start with the cdf.

$$F_Y(x) = P(Y \leq x) = P\left(|X| \leq x\right).$$

Now stop, to analyze the situation. Since $|X| \geq 0$, it is *impossible* for $|X|$ to be less than or equal to a negative number, so for $x < 0$, $F_Y(x) = 0$ and, hence,

$$f_Y(x) = F_Y'(x) = 0,\ x < 0.$$

For $x \geq 0$, we rewrite the inequality until *it is an inequality for* X, whose pdf is *known*, so probabilities about X can be computed. We have

$$F_Y(x) \;=\; P\left(-x \leq X \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

Now, the function $e^{-\frac{t^2}{2}}$ has no primitive, so what do we do, how do we compute that integral? **We don't!** Remember what we want is the *pdf*, i.e. the *derivative of* F, so we want to go with the computations just far enough to be able to find its derivative. Even if we could compute it, there's no need to! All we want is to write it in the form

$$G(x) \;=\; \int_{a}^{x} g(t)\, dt,\; a \in \mathbb{R},$$

for which the derivative is

$$G'(x) \;=\; g(x).$$

Or, more generally,

$$G(x) \;=\; \int_{a}^{h(x)} g(t)\, dt,\; a \in \mathbb{R} \;<=>\; G'(x) \;=\; g\big(h(x)\big) \cdot h'(x),$$

which also involves the derivative of a composite function.
So, to get back to our integral, let us recall another thing about integration and symmetry:

$$\int_{-a}^{a} f(x)\, dx \;=\; \begin{cases} 2\displaystyle\int_{0}^{a} f(x)\, dx, & \text{if the function is even, i.e. } f(-x) = f(x) \\ 0, & \text{if the function is odd, i.e. } f(-x) = -f(x) \end{cases}.$$

Since we are integrating over a symmetric interval and the function $e^{-\frac{1}{2}t^2}$ is even, we have

$$F_Y(x) \;=\; 2\frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{1}{2}t^2}\, dt \;=\; \sqrt{\frac{2}{\pi}} \int_{0}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

Now we can take the derivative

$$F_Y'(x) \;=\; \sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}x^2}.$$

Thus, the pdf of Y is

$$f_Y(x) \;=\; \sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}x^2},\; x \geq 0 \;\text{ (and } 0, \text{ otherwise).}$$