1 Continuous Random Variables

The definition of the CDF is identical for all random variables whether they are discrete or not.

Definition 1 (Cumulative Distribution Function)**.** The *cumulative distribution function* (CDF) of a random variable X is

$$F_X(x) = \mathbb{P}(X \leq x) := \mathbb{P}(\{\omega \in S : X(\omega) \leq x\}), \quad x \in \mathbb{R}.$$

From the CDF, we can easily compute any probability, since the intervals $(a, b]$ and $(-\infty, a]$ are disjoint,

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F_X(b) - F_X(a), \tag{1}$$

so the CDFs encode the same information as a probability distribution. From the point of view of CDFs, we have the following natural classifications of random variables.

Definition 2 (Discrete vs Continuous Random Variable)**.** If the CDF of X is

1. a *piecewise constant* function, then X is a *discrete* random variable.
2. a *continuous* function, then X is a *continuous* random variable.

If the CDF is a continuous function, then it immediately follows from (1) that $\mathbb{P}(X = x) = 0$, so

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b).$$

Remark 1. It is possible that a CDF does not fall under either of these categories, such as mixed random variables with have CDFs with both jump discontinuities and strictly increasing parts.

1.1 Probability Density Function (PDF)

We can try to define the PMF for a continuous random variable, but the fact that $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$ is problematic. We get around this by defining a new way to encode how likely a certain value x is without defining it as a probability.

Definition 3 (Absolutely Continuous Random Variable)**.** We say that a continuous random variable X with distribution function F_X is *absolutely continuous* if it is the antiderivative of some function f_X,

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

Remark 2. It might be the case that $F_X(x)$ is not differentiable at isolated points. However, if $f_X(x)$ is continuous at x, then

$$F_X'(x) = \frac{d}{dx} F_X(x) = f_X(x).$$

Likewise, jump discontinuities of f_X correspond to non-differentiable points of F_X.

Definition 4 (Probability Density Function)**.** The function f_X is called the *probability density function* (PDF). The function $f_X(x)$ and it satisfies the following properties

1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$;
2. $\int_{-\infty}^{\infty} f_X(x) dx = 1$;

The *support* of f_X (or X) is the closure of the set of non-zero values of the PDF,

$$\mathrm{supp}(f_X) = \mathrm{cl}(\{x \in \mathbb{R} : f_X(x) \neq 0\}).$$

Remark 3. By convention we take the closure of the set, which means that we always include the endpoints of intervals in the support. This is not a big issue since $f_X(x)$ can be arbitrarily defined at the endpoints since the area under the PDF does not change if we redefine the endpoints.

The PDF $f_X(x)$ is proportional to the probability that X lies in a small interval around x in the sense that for ϵ small

$$\mathbb{P}\left(x - \frac{\epsilon}{2} \leq X \leq x + \frac{\epsilon}{2}\right) = \int_{x-\frac{\epsilon}{2}}^{x+\frac{\epsilon}{2}} f_X(x)\, dt \approx \epsilon f_X(x).$$

So $f_X(x)$ isn't the probability that $X = x$, but it encodes the likelihood of x compared to other values. In fact, $f_X(x)$ cannot be a probability since it is possible for it to take values larger than 1.

Remark 4. In this course, unless otherwise stated, all continuous random variables will be absolutely continuous so we often drop the prefix "absolutely".

1.2 Change of Variables Formula

Suppose that we know the PDF $f_X(x)$ of X. Our goal is to recover the the PDF $f_Y(y)$ of the random variable $Y = g(X)$. This can be done directly using the following steps

1. Use the support of X to find the support of $Y = g(X)$:

$$\text{supp}(Y) = \text{cl}(\{y \in \mathbb{R} : f_Y(y) > 0\}) = g(\text{supp}(X)).$$

2. Compute the CDF of Y for $y \in \text{supp}(Y)$ by expressing it in terms of the CDF of X:

$$F_Y(y) = \mathbb{P}(g(X) \leq y) = \dots$$

   When the function g is not strictly increasing (or decreasing) over the support of X, then we must be careful when rewriting the inequality $\mathbb{P}(g(X) \leq y)$.

3. Compute the PDF of Y by differentiating the CDF of Y,

$$f_Y(y) = F_Y'(y) \qquad y \in \text{supp}(Y).$$

Remark 5. Technically, the function F_Y is only differentiable on the interior of $\text{supp}(Y)$. This is not an issue since we can simply define $f_Y(y)$ on the boundaries by continuity.

When g is invertible, the above procedure gives us the change of density formula.

Theorem 1 (*Change of Variables Formula*)

Let X be a (absolutely) continuous random variable and g be invertible and differentiable with inverse g^{-1} on the support of Y, then

$$f_Y(y) = |(g^{-1})'(y)| f_X(g^{-1}(y)) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)), \quad y \in \text{supp}(Y).$$

1.3 Expected Value and Variance

The expected value and variance of a random variable can be defined analogously to the discrete random variables where the sum is now replaced by an integral.

Definition 5 (Law of the Unconscious Statistician)**.** If X is a continuous random variable with PDF $f_X(x)$, and $g : \mathbb{R} \to \mathbb{R}$ is a function, then

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

provided the expression exists.

It follows that for continuous random variables,

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

and

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x) dx.$$

This is analogous to how the expected value and variance of discrete random variables were defined, but the sum over the PMF is replaced with an integral over the PDF.

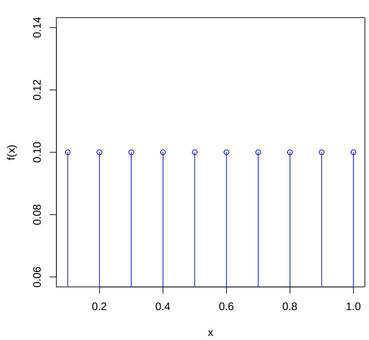
1.4 Example Problems

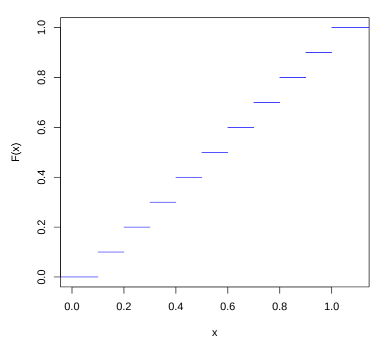
Problem 1.1. Suppose we are cutting a stick of length 1 randomly and denote by X the cutting point. The random variable X is continuous with range $[0, 1]$.

1. Approximate X with a discrete uniform distribution on $\{\frac{1}{10}, \frac{2}{10} \dots, 1\}$. Draw the PMF and CDF.
2. Approximate X with a discrete uniform distribution on $\{\frac{1}{20}, \frac{2}{20} \dots, 1\}$. Draw the PMF and CDF.
3. Approximate X with a discrete uniform distribution on $\{\frac{1}{100}, \frac{2}{100} \dots, 1\}$. Draw the PMF and CDF.
4. Approximate X with a discrete uniform distribution on $\{\frac{1}{10000}, \frac{2}{10000} \dots, 1\}$. Draw the PMF and CDF.

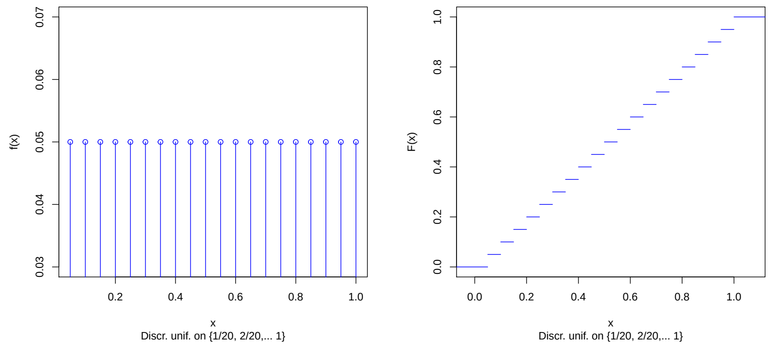
Solution 1.1.

1. We have $f_X(x) = \frac{1}{10}$ for all $x \in \{\frac{1}{10}, \frac{2}{10} \dots, 1\}$.

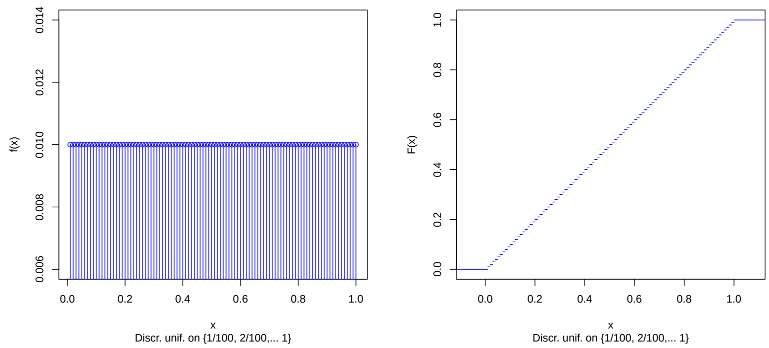




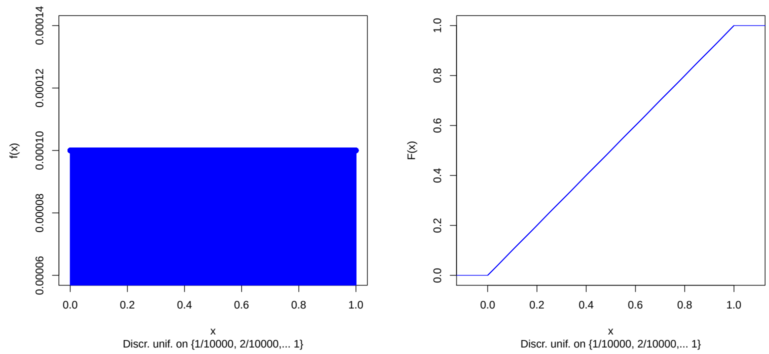
2. We have $f_X(x) = \frac{1}{20}$ for all $\{\frac{1}{20}, \frac{2}{20} \dots, 1\}$.



3. We have $f_X(x) = \frac{1}{100}$ for all $x \in \{\frac{1}{100}, \frac{2}{100} \dots, 1\}$



4. We have $f_X(x) = \frac{1}{10000}$ for all $x \in \{\frac{1}{10000}, \frac{2}{10000} \dots, 1\}$.



Remark 6. Notice that the PMF gradually tends to 0 and the CDF gets closer to CDF of the continuous uniform distribution on $[0, 1]$.

Problem 1.2. Suppose that X is a continuous random variable with PDF

$$f_X(x) = \begin{cases} cx(1-x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

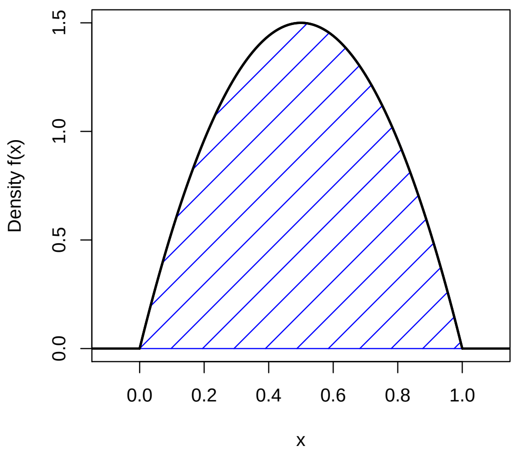
1. Compute c so that this is a valid pdf.

2. Compute the cdf $F_X(x)$.
3. Compute $\mathbb{P}(1/4 \le X \le 3/4)$

Solution 1.2.

1. We need $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$, so

$$\int_0^1 cx(1-x)\,dx = \frac{c}{6} = 1 \implies c = 6.$$

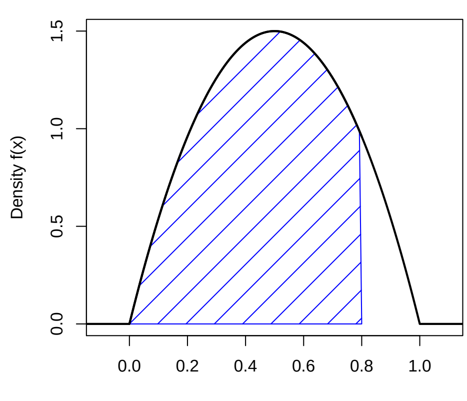


Blue area = 1

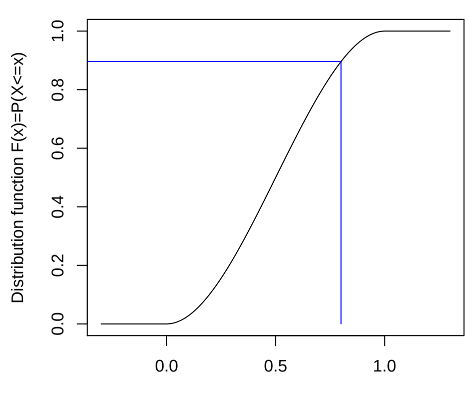
2. The CDF $F_X(x)$ is

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x 6t(1-t)\,dt = 3x^2 - 2x^3, \quad x \in [0,1]$$

and $F_X(x) = 0$ for $x \le 0$ and $F_X(x) = 1$ for $x \ge 1$.

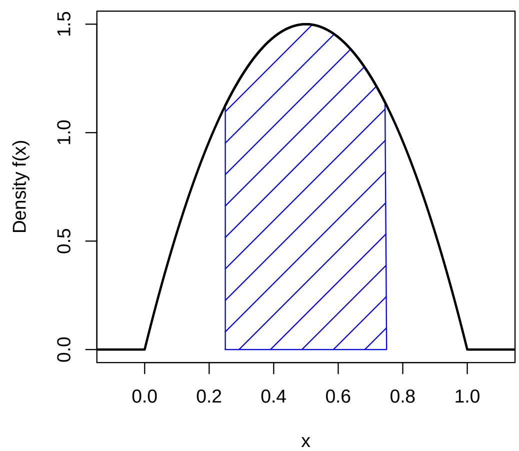


Blue area = P(X <= 0.8) = F(0.8)



3. We compute the probability using the CDF:

$$\mathbb{P}\left(\frac{1}{4} \le X \le \frac{3}{4}\right) \overset{\text{cont.}}{=} \mathbb{P}\left(\frac{1}{4} < X \le \frac{3}{4}\right) = \mathbb{P}(X \le 3/4) - P(X \le 1/4)$$
$$= F_X(3/4) - F_X(1/4) = 11/16.$$



Blue area = P(0.25 < X <= 0.75)

Alternative Solution: By integrating the PDF

$$\mathbb{P}(1/4 \le X \le 3/4) = \int_{1/4}^{3/4} f_X(t)\,dt = 11/16.$$

Problem 1.3. Suppose the random variable X has PDF

$$f_X(x) = \begin{cases} 6x(1-x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

Compute $\mathbb{E}(X)$ and $\mathrm{Var}(X)$.

Solution 1.3. By definition,

$$\mathbb{E}(X) = \int_0^1 x \cdot 6x(1-x)\,dx = \frac{1}{2}$$

and

$$\mathbb{E}(X^2) = \int_0^1 x^2 \cdot 6x(1-x)\,dx = \frac{3}{10}.$$

Therefore,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{3}{10} - \left(\frac{1}{2}\right)^2 = \frac{1}{20} = 0.05$$

Problem 1.4. Suppose X has PDF $f_X(x)$, and f_X is an even function about the origin on $\mathbb{R}$ (i.e. $f_X(x) = f_X(-x)$). If $\mathbb{E}[X]$ is well defined, show that $\mathbb{E}[X] = 0$.

Solution 1.4. Since $f_X(x) = f_X(-x)$, the "positive and negative areas" cancel since $x \cdot f_X(x)$ is an odd function. To see this directly, notice that

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx \\
&= \int_{-\infty}^{0} x \cdot f_X(x)\, dx + \int_{0}^{\infty} x \cdot f_X(x)\, dx \\
&= \int_{0}^{\infty} (-x) \underbrace{f_X(-x)}_{=f_X(x)}\, dx + \int_{0}^{\infty} x \cdot f_X(x)\, dx \\
&= -\int_{0}^{\infty} x \cdot f_X(x)\, dx + \int_{0}^{\infty} x \cdot f_X(x)\, dx \\
&= 0
\end{aligned}$$

Problem 1.5. Let X be a continuous random variable with CDF $F_X(x) = \mathbb{P}(X \le x)$ and let $g : \mathbb{R} \mapsto \mathbb{R}$ be an increasing function with inverse g^{-1}. Compute $F_Y(y) = \mathbb{P}(Y \le y)$.

Solution 1.5. We can write the CDF of Y in terms of the CDF of X by

$$F_Y(y) = \mathbb{P}(g(X) \le y) = \mathbb{P}(X \le g^{-1}(y)) = F_X(g^{-1}(y)).$$

Remark 7. Notice that the CDF of $Y = g(X)$ is not $g(F_X(x))$.

Problem 1.6. Let X be a continuous random variable with the following pdf:

$$f_X(x) = \begin{cases} \frac{1}{4} & 0 < x \le 4, \\ 0 & \text{otherwise} \end{cases}$$

1. Find the CDF of X.
2. Let $Y = X^{-1}$. Find the CDF Y.
3. Find the PDF of Y.

Solution 1.6.

1. Outside the support of X, we have $F_X(x) = 0$ for $x < 0$ and $F_X(x) = 1$ for $x > 4$. For x in the support of X, i.e. $x \in [0, 4]$
$$F_X(x) = \int_0^x \frac{1}{4} = \frac{x}{4}.$$
In summary,
$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{x}{4}, & \text{if } 0 \le x < 4 \\ 1, & \text{if } 4 \le x \end{cases}$$

2. Notice that $\text{supp}(X) = [0,4]$. Therefore, the support of $Y = X^{-1}$ is

$$0 \le X \le 4 \implies \infty \ge X^{-1} \ge \frac{1}{4} \implies \text{supp}(Y) = \Big[\frac{1}{4}, \infty\Big).$$

For y in the support of Y, i.e. $y \in \Big[\frac{1}{4}, \infty\Big)$

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(X^{-1} \le y) = \mathbb{P}(X > \frac{1}{y}) = 1 - \mathbb{P}(X \le \frac{1}{y}) = 1 - F_X(y^{-1}) = 1 - \frac{1}{4y}.$$

In summary,

$$F_Y(y) = \begin{cases} 0, & \text{if } y < \frac{1}{4} \\ 1 - \frac{1}{4y}, & \text{if } \frac{1}{4} \le y \end{cases}$$

3. To get the PDF of Y, we simply differentiate $F_Y(y)$ to conclude that

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\Big(1 - \frac{1}{4y}\Big) = \frac{1}{4y^2} \qquad \text{for } y \in \Big[\frac{1}{4}, \infty\Big).$$

and $f_Y(y) = 0$ for $y \notin \Big[\frac{1}{4}, \infty\Big)$.

Alternative Solution: We can compute the PDF of f using the change of variables formula. We have $g(x) = \frac{1}{x}$, so $g'(x) = -\frac{1}{x^2}$ and $g^{-1}(x) = \frac{1}{x}$. Therefore, for $y \in [\frac{1}{4}, \infty)$,

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)) = \frac{1}{|-\frac{1}{(y^{-1})^2}|}\frac{1}{4} = \frac{1}{4y^2}.$$

Problem 1.7. Suppose a continuous random variable X has probability density function

$$f_X(x) = \begin{cases} 1 - |x| & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Find the CDF of X.
2. Let $Y = X^2$. Find the CDF Y.
3. Find the PDF of Y.

Solution 1.7.

1. We can rewrite the PDF as

$$f_X(x) = \begin{cases} 1 + x & -1 \le x < 0 \\ 1 - x & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Outside the support of X, we have $F_X(x) = 0$ for $x < -1$ and $F_X(x) = 1$ for $x > 1$. We have two cases for x in the support of X,

(a) If $x \in [-1, 0]$

$$F_X(x) = \int_{-1}^{x} 1 + t\,dt = t + \frac{t^2}{2}\Big|_{-1}^{x} = x + \frac{x^2}{2} + \frac{1}{2}.$$

(b) If $x \in [0, 1]$

$$F_X(x) = \int_{-1}^{0} 1 + t\,dt + \int_{0}^{x} 1 - t\,dt = \Big(t + \frac{t^2}{2}\Big)\Big|_{-1}^{0} + \Big(t - \frac{t^2}{2}\Big)\Big|_{0}^{x} = x - \frac{x^2}{2} + \frac{1}{2}.$$

In summary,

$$F_X(x) = \begin{cases} 0 & x < -1 \\ x + \frac{1}{2}x^2 + \frac{1}{2} & -1 \le x < 0 \\ x - \frac{1}{2}x^2 + \frac{1}{2} & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$

2. Notice that $\mathrm{supp}(X) = [-1, 1]$. Therefore, the support of $Y = X^2$ is

$$-1 \le X \le 1 \implies 0 \le X^2 \le 1 \implies \mathrm{supp}(Y) = [0, 1].$$

Outside of the support, for $y < 0$ we have $F_Y(y) = 0$ and for $y > 1$ we have $F_Y(y) = 1$. In the support of Y, we have

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= \left(\sqrt{y} - \frac{1}{2}y + \frac{1}{2}\right) - \left(-\sqrt{y} + \frac{1}{2}y + \frac{1}{2}\right) \\ &= 2\sqrt{y} - y. \end{aligned}$$

In summary,

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ 2\sqrt{y} - y & 0 \le y < 1 \\ 1 & 1 \le y \end{cases}$$

3. To get the PDF of Y, we simply differentiate $F_Y(y)$ to conclude that

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\Big(2\sqrt{y} - y\Big) = \frac{1}{\sqrt{y}} - 1 \qquad \text{for } y \in [0, 1].$$

and $f_Y(y) = 0$ for $y \notin [0, 1]$.

Remark 8. The function $g(x) = x^2$ is not increasing on the $\mathrm{supp}(X) = [-1, 1]$ so we can't use the change of variables formula.

1.5 Proofs of Key Results

Problem 1.8. Suppose X is a continuous random variable. Show that

$$\mathbb{P}(X = x) = 0$$

and

$$\mathbb{P}(a < X \le b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \le X < b) = \mathbb{P}(a \le X \le b) = F_X(b) - F_X(a).$$

Solution 1.8. If X is a continuous random variable, then for any $x \in \mathbb{R}$,

$$\mathbb{P}(X = x) = \lim_{\epsilon \to 0} \mathbb{P}(x - \epsilon < X \le x) = F_X(x) - F_X(x^-) = 0$$

by continuity of F_X. This implies that the inequalities don't matter for continuous random variables

$$\mathbb{P}(a < X \le b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \le X < b) = \mathbb{P}(a \le X \le b) = F_X(b) - F_X(a).$$

Remark 9. If X is discrete, then the inequalities matter, so

$$\mathbb{P}(a < X \le b),\ \mathbb{P}(a < X < b),\ \mathbb{P}(a \le X < b),\ \mathbb{P}(a \le X \le b)$$

can be different since $\mathbb{P}(X = a)$ or $\mathbb{P}(X = b)$ may be non-zero.

Problem 1.9. Prove the change of variables formula in Theorem 1.

Solution 1.9.

Strictly Increasing: We first consider the case that g is strictly increasing. If g is strictly increasing, then

$$F_Y(y) = \mathbb{P}(g(X) \le y) = \mathbb{P}(X \le g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(t)\,dt.$$

Therefore, we can use the fundamental theorem of calculus and the chain rule to see that for points in the interior of the support of Y,

$$f_Y(y) = \frac{d}{dy}\int_{-\infty}^{g^{-1}(y)} f_X(t)\,dt = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) = \frac{1}{g'(g^{-1}(y))} f_X(g^{-1}(y)).$$

Since g is strictly increasing, $g' > 0$, so

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)), \quad y \in \mathrm{supp}(Y).$$

Strictly Decreasing: We now consider the case that g is strictly decreasing. If g is strictly decreasing, then

$$F_Y(y) = \mathbb{P}(g(X) \le y) = \mathbb{P}(X \ge g^{-1}(y)) = 1 - \mathbb{P}(X \le g^{-1}(y)) = 1 - 1 - \int_{-\infty}^{g^{-1}(y)} f_X(t)\,dt.$$

Therefore, we can use the fundamental theorem of calculus and the chain rule to see that for points in the interior of the support of Y,

$$f_Y(y) = \frac{d}{dy}\left(1 - \int_{-\infty}^{g^{-1}(y)} f_X(t)\,dt\right) = -f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) = -\frac{1}{g'(g^{-1}(y))} f_X(g^{-1}(y)).$$

Since g is strictly increasing, $g' < 0$, so

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)), \quad y \in \mathrm{supp}(Y).$$

2 Important Continuous Random Variables

2.1 Uniform Distribution: $\mathrm{Unif}(a, b)$

The uniform distribution models variables with equally likely outcomes on an interval. This is the continuous analogue of the discrete uniform distribution.

Definition 6 (Uniform Distribution)**.** We say that X has a *continuous uniform distribution* on (a, b) if X has PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b], \\ 0 & \text{otherwise} \end{cases}$$

This is denoted by

$$X \sim \mathrm{Unif}(a, b).$$

Example 1. The following experiments can be modeled by a continuous uniform distribution:

Experiment	X	Distribution
Cutting a stick of length 1 at a random position	the location of the cut	$\mathrm{Unif}(0, 1)$
Spinning a wheel	Location of spinner	$\mathrm{Unif}(0, 2\pi)$

2.1.1 Properties

1. Sampling uniformly on the intervals (a, b), $[a, b)$, $(a, b]$ and $[a, b]$ are all equivalent.
2. **Mean and Variance:** If $X \sim \mathrm{U}(a, b)$ then

$$\mathbb{E}[X] = \frac{a+b}{2} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

2.2 Exponential Distribution: $\mathrm{Exp}(\theta)$

The exponential distribution models the time between occurrences of a Poisson process with rate parameter $\lambda = \frac{1}{\theta}$ expressed in occurrences per time. The mean parameter θ is expressed in time per occurrence. This is the continuous time analogue of the geometric distribution.

Definition 7 (Exponential Distribution)**.** We say that X has an exponential distribution with *mean* parameter θ if X has PDF

$$f_X(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x \geq 0, \\ 0 & x < 0 \end{cases}$$

This is denoted by $X \sim \mathrm{Exp}(\theta)$.

Example 2. The following experiments can be modeled by a exponential distribution:

Experiment	X	Distribution
Busses arriving at rate 3 per hour	hours until a bus arrives	$\mathrm{Exp}(3^{-1})$
Busses arriving every 15 minutes	minutes until a bus arrives	$\mathrm{Exp}(15)$
Emails arrive at a rate of 20 per hour	hours until the next email	$\mathrm{Exp}(20^{-1})$

Remark 10. We can also parameterize the exponential random variable with the rate parameter. We say that X has an exponential distribution with rate parameter λ if X has PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & x \leq 0 \end{cases}$$

This is denoted by $X \sim \mathrm{Expo}(\lambda)$ in the textbook. Unless otherwise stated, we will take the mean parametrization as the standard one in this course since that is the one given in your formula sheet.

2.2.1 Poisson Process

We now introduce a stochastic process such that the number of points in any time interval is given by a Poisson random variable. It is used to model random events that happen at a consistent rate and generalizes the Poisson random variable to variable times.

Definition 8 (Poisson Process)**.** A stochastic process $\{N(t)\}_{t\geq 0}$ is called a **Poisson process** with rate $\lambda > 0$ if

(1) $N(0) = 0$.

(2) It has independent increments. i.e., for any $0 \leq t_0 < t_1 < \cdots < t_n$, the increments

$$N(t_1) - N(t_0),\ N(t_2) - N(t_1), \dots,\ N(t_n) - N(t_{n-1})$$

are all independent.

(3) For any $t \geq 0$ and $h > 0$, the increment $N(t+h) - N(t)$ has a Poisson distribution with parameter λh, i.e.,

$$\mathbb{P}\left(N(t+h) - N(t) = n\right) = e^{-\lambda h} \frac{(\lambda h)^n}{n!}, \quad n = 0, 1, 2, \dots$$

On any time interval of length h, the number of points has a Poisson distribution with mean λh, so the number of occurrences is proportional to the length of time.

2.2.2 Properties

1. **Parameters:** If a Poisson process has rate λ occurrences per time, then $\theta = \frac{1}{\lambda}$ is the mean time per occurrence.

2. **Mean and Variance:** If $X \sim \text{Exp}(\theta)$ then

$$\mathbb{E}[X] = \theta \qquad \text{Var}(X) = \theta^2.$$

3. **Memoryless Property:** The exponential distribution forgets how long we have waited already.

Theorem 2 (*Memoryless property of Exp(θ)*)

If $X \sim \text{Exp}(\theta)$, then

$$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t).$$

In fact, the exponential distribution is the only continuous random variable with this property.

4. Computations with the exponential distribution can be expressed using the Gamma function

Definition 9 (Gamma function)**.** The integral

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha - 1} e^{-y} dy,\ \alpha > 0$$

is called the gamma function of α.

The Gamma function is a continuous function that interpolates the factorial function and it satisfies the following nice properties

(a) $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$
(b) $\Gamma(1/2) = \sqrt{\pi}$
(c) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for $\alpha > 1$

2.3 Normal Distribution: $\mathrm{N}(\mu, \sigma^2)$

The normal distribution, one of the most widely and most important continuous distributions in theory and applications. A lot of data is unimodal, symmetric around the mean μ, and the majority of data is "near μ" and few are "far from μ".

Definition 10 (Gaussian Distribution)**.** We say that X has a *normal distribution* (or Gaussian distribution) with mean μ and variance σ^2 if X has PDF

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

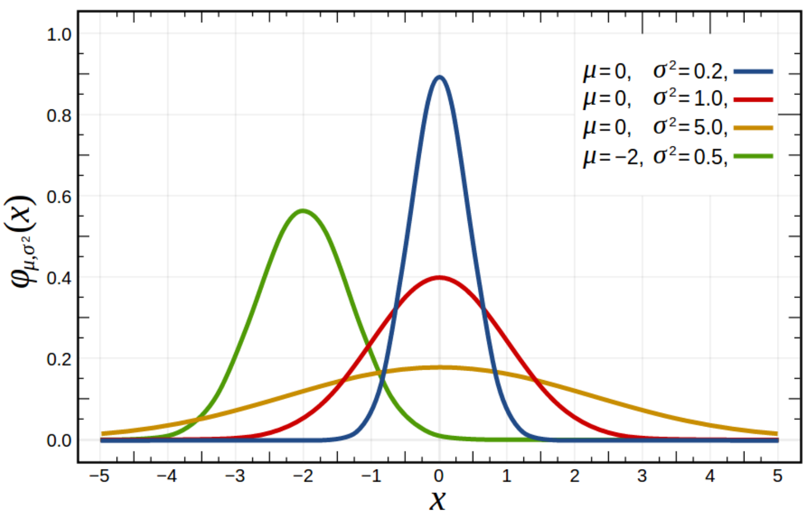
This is denoted by

$$X \sim \mathrm{N}(\mu, \sigma^2).$$

Example 3. The following experiments can be modeled by a continuous uniform distribution:

Experiment	X	Distribution
Exams with 70% average and standard dev 20%	the grade of a random student	$\mathrm{N}(0.7, 0.2^2)$
IQ test (average IQ of 100 and 15 standard dev)	IQ of a random person	$\mathrm{N}(100, 15^2)$

Example 4. Several PDFs of normally distributed random variables are below. The mean is where the peak of the PDF is, and the variance encodes how spread out the PDF is.



2.3.1 Properties

1. **Mean and Variance:** If $X \sim \mathrm{N}(\mu, \sigma^2)$ then
$$\mathbb{E}[X] = \mu \qquad \mathrm{Var}(X) = \sigma^2.$$

2. **Symmetric about its mean:** If $X \sim \mathrm{N}(\mu, \sigma^2)$
$$\mathbb{P}(X \leq \mu - t) = \mathbb{P}(X \geq \mu + t).$$

3. **Density is unimodal:** Peak is at μ.

4. **68-95-99.7 Rule:** If $X \sim \mathrm{N}(\mu, \sigma^2)$, then
$$\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68 \qquad \text{68\% values lie within 1 standard deviation of the mean}$$
$$\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95 \qquad \text{95\% values lie within 2 standard deviation of the mean}$$
$$\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997 \qquad \text{99.7\% values lie within 3 standard deviation of the mean .}$$

2.3.2 Standard Normal Distribution: $\mathrm{N}(0,1)$

Definition 11 (Standard Gaussian)**.** We say that Z follows the *standard normal distribution* if $Z \sim \mathrm{N}(0,1)$. The PDF of the standard normal random variable is denoted

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}},$$

and the CDF of a standard normal random variable is denoted

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} \, dy.$$

Every normally distributed random can be transformed into a normally distributed random variables through a linear transformation. This technique is called standardization.

Theorem 3 (*Standardizing normal random variables*)

If $X \sim \mathrm{N}(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim \mathrm{N}(0,1).$$

Remark 11. We use the transformation $Z = \frac{X-\mu}{\sigma}$ to go from X to a standard normal. And we use the transformation $X = \mu + \sigma Z$ to go from a standard normal to X.

2.4 Example Problems

Problem 2.1. Suppose $X \sim \mathrm{Unif}(0,1)$, and that $Y = \frac{2}{X} - 1$. What is the support of Y?

Solution 2.1. Since $\mathrm{supp}(X) = [0,1]$, we have

$$0 \leq X \leq 1 \implies \infty > \frac{2}{X} \geq 2 \implies \infty > \frac{2}{X} - 1 \geq 1$$

so $\mathrm{supp}(Y) = [1, \infty)$.

Problem 2.2. Suppose that the angle measured from the principal axis to the point of a spinner is uniformly distributed on $[0, 2\pi]$. You win the prize you want if the point lands in $[\frac{3\pi}{4}, \frac{3\pi}{2}]$. Given that the point will stop in the bottom half of the circle, what is the probability that you win the prize you want.

Solution 2.2. If X denotes the angle of the spinning wheel, then $X \sim U(0, 2\pi)$. Note that the bottom half is the circle $[\pi, 2\pi]$. Denote by A_1, A_2 the events $A_1 = \{X \in [\pi, 2\pi]\}$ and $A_2 = \{X \in [\frac{3\pi}{4}, \frac{3\pi}{2}]\}$. The desired probability is

$$\begin{aligned}
\mathbb{P}(A_2 \mid A_1) = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} &= \frac{\mathbb{P}(X \in [\pi, 2\pi] \cap [\frac{3\pi}{4}, \frac{3\pi}{2}])}{\mathbb{P}(X \in [\pi, 2\pi])} \\
&= \frac{\mathbb{P}(X \in [\pi, \frac{3\pi}{2}])}{\mathbb{P}(X \in [\pi, 2\pi])} \\
&= \frac{\frac{1}{2\pi} \cdot (\frac{3\pi}{2} - \pi)}{\frac{1}{2\pi} \cdot (2\pi - \pi)} \\
&= \frac{1/4}{1/2} = 1/2.
\end{aligned}$$

Problem 2.3. Nupur decided to enjoy a relaxing summer away from student housing, so she rented a place in Simcoe, Ontario. However, the busses there are far and few between. Suppose busses arrive according to a Poisson process with an average of 3 busses per hour.

1. Find the probability of waiting at least 15 minutes.
2. Find the probability of waiting at least another 15 minutes given that you have already been waiting for 6 minutes.

Solution 2.3.

Part 1: If X denotes the waiting time until the first bus in minutes, then the waiting time parameter is $\theta = 20$ min per bus so $X \sim \text{Exp}(20)$, and

$$\mathbb{P}(X \geq 15) = \int_{15}^{\infty} \frac{1}{20} e^{-\frac{1}{20}x} dx = e^{-\frac{3}{60}\cdot 15} = e^{-3/4}$$

Alternative Solution: If X denotes the waiting time until the first bus in hours, then the waiting time parameter is $\theta = \frac{1}{3}$ hours per bus (rate λ is 3 busses per hour) $X \sim \text{Exp}(3^{-1})$. Hence, the probability of waiting at least 15 minutes is

$$\mathbb{P}(X \geq 1/4) = \int_{1/4}^{\infty} 3e^{-3x} dx = e^{-3\cdot\frac{1}{4}}.$$

Part 2: By the memoryless property, we have

$$\mathbb{P}(X \geq 15 + 6 \mid X \geq 6) = \mathbb{P}(X \geq 15) = e^{-3/4}$$

from above.

Problem 2.4. Exponential distribution is also very useful in reliability engineering. The lifetime of a seat belt motor on a 1994 Saturn GL is known to follow an exponential distribution with mean 14 years.

1. What is the standard deviation of the lifetime of a seat belt motor on a 1994 Saturn GL?
2. Compute the probability that the lifetime of the seat belt motor will last more than 6 years.
3. If a seat belt motor has lasted 14 years, what is the probability that it will last another 6 years?

Solution 2.4. In this problem we are given $\mathbb{E}[X] = 14$ and X has exponential distribution so $X \sim \text{Exp}(14)$.

1. We have $\text{Var}(X) = 14^2$, so the standard deviation is 14 years.
2. We have
$$\mathbb{P}(X \geq 6) = \int_{6}^{\infty} \frac{1}{14} e^{-\frac{1}{14}t}\, dt = e^{-\frac{6}{14}}.$$
3. By the memoryless property, we have
$$\mathbb{P}(X \geq 6 + 14 \mid X \geq 14) = \mathbb{P}(X \geq 6) = e^{-\frac{6}{14}}.$$

Problem 2.5. Suppose the waiting time X until Mukhtar's next bus arrives follows an exponential distribution with parameter $\theta = 1$. What's the waiting time w so that Mukhtar doesn't have to wait longer than w with probability 50%?

Solution 2.5. We know

$$F_X(x) = \mathbb{P}(X \le x) = \int_0^x e^{-t}\, dt = 1 - e^{-x}$$

for $x \ge 0$ and 0 otherwise. This function is strictly increasing on $x > 0$, so we can use the classical inverse. We want w such that $\mathbb{P}(X \le w) = 0.5$, or

$$F_X(w) = 0.5 \Leftrightarrow 1 - e^{-w} = 0.5 \Leftrightarrow w = \log(2) \approx 0.693.$$

So with probability 50% Mukhtar won't have to wait longer than $\log(2)$. In other words, $\log(2)$ is the median of the distribution of X.

Remark 12. We can repeat this computation for general $\theta \neq 1$. The median in this case will be given by $\theta \log(2)$. Since $\log(2) < 1$ this implies that the median of the exponential is always below the mean.

Problem 2.6. Uranium 238 emits particles measured by a Geiger counter at a rate of 50 per second. Assume that the number of particles measured by a Geiger counter follows a Poisson process. Let X denote the amount of time in seconds between when the first and second particles are measured. Find $\mathbb{E}[X]$

Solution 2.6. Intuitively, since the rates are homogeneous and the increments are independent, the time between the first and second particle is equal in distribution to the time between the first particle. Therefore, the waiting time parameter θ is $\frac{1}{50}$ seconds per occurrence, so $X \sim \mathrm{Exp}(50^{-1})$. Therefore, $\mathbb{E}[X] = \theta = \frac{1}{50}$.

We carefully explain how the homogeneous and independence property are used in this problem. Let X be the time of the first particle, Y be the time from the first to the second particle, and $N(t)$ be the number of particles by time t. Clearly, $Z = X + Y$ is the time of the second particle.

We know $X \sim \mathrm{Exp}(50^{-1})$ and $N(t) \sim \mathrm{Poi}(50t)$. Using the same logic as in the derivation of the Poisson process (see Problem (2.10)),

$$F_Z(t) = \mathbb{P}(X + Y \le t) = 1 - \mathbb{P}(X + Y > t) = 1 - P(N(t) \le 1) = 1 - (e^{-50t} + e^{-50t}(50t))$$

So $Z = X + Y$ has density

$$f_Z(t) = \frac{d}{dt}\mathbb{P}(X + Y \le t) = -\frac{d}{dt}\left[e^{-50t}(1 + 50t)\right] = t50^2 e^{-50t}$$

Therefore, by an integration by parts

$$\mathbb{E}[Z] = \int_{-\infty}^{\infty} z \cdot f_Z(z) dz = 50 \int_0^{\infty} z^2 \cdot 50 e^{-50z} dz = 50 \frac{2}{50^2} = \frac{2}{50}$$

and the linearity of expectation implies that

$$\mathbb{E}[Y] = \mathbb{E}[Z - X] = \mathbb{E}[Z] - \mathbb{E}[X] = \frac{2}{50} - \frac{1}{50} = \frac{1}{50}.$$

Problem 2.7. At a super busy coffee chain, customers arrive according to a Poisson process at a rate of $\lambda = 5$ customers per minute.

1. Find the probability that there are more than 2 customers in one minute.
2. Suppose you record the number of customers in 5 consecutive one-minute intervals. What is the probability that in at least 3 of them there were more than 2 customers?
3. Find the probability that a minute with more than 2 customers actually had 6 customers
4. Suppose you are waiting until finally, there is one minute with more than 2 customers. Denote by X the the number of minutes you need to wait. Find the PMF of X.
5. Suppose in 3 minutes, there were n customers. Find the probability that x of these came in the first two minutes.

Solution 2.7. Let t be measured in minutes, and the rate is $\lambda = 5$ customers per minute.

1. If X is the number of customers in one minute, then $X \sim \mathrm{Poi}(5 \cdot 1)$. Thus,
$$\begin{aligned} p = \mathbb{P}(X > 2) &= 1 - \mathbb{P}(X \leq 2) \\ &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) \\ &= 1 - e^{-5}\Big(1 + 5 + \frac{5^2}{2}\Big) \approx 0.875 \end{aligned}$$

2. Let Y be the number of one-minute intervals with more than two customers, then $Y \sim \mathrm{Bin}(5, p)$ with p from earlier. Thus,
$$\begin{aligned} \mathbb{P}(Y \geq 3) &= \mathbb{P}(Y = 3) + \mathbb{P}(Y = 4) + \mathbb{P}(Y = 5) \\ &= \binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p) + \binom{5}{5} p^5 (1-p)^0 \approx 0.984 \end{aligned}$$

3. Let X be the number of customers in one minute. Thus,
$$\mathbb{P}(X = 6 \mid X > 2) = \frac{\mathbb{P}(X = 6 \text{ and } X > 2)}{\mathbb{P}(X > 2)} = \frac{\mathbb{P}(X = 6)}{\mathbb{P}(X > 2)} = \frac{e^{-5}\frac{5^6}{6!}}{0.875} \approx 0.167$$

4. Let Z be the number of minutes until first minute with more than 2 customers, then $Z \sim Geo(p)$ with p from earlier. Thus,
$$f_Z(x) = \mathbb{P}(Z = x) = (1-p)^x p, \quad x = 0, 1, 2, \ldots$$

5. We want to fine
$$\mathbb{P}(x \text{ in first 2min} \mid n \text{ in 3min}) = \frac{\mathbb{P}(x \text{ in first 2min and n in 3min})}{\mathbb{P}(n \text{ in 3min})}.$$
We know that
   - Denominator: The number of customers in 3 minutes follows a $\mathrm{Poi}(5 \cdot 3)$ distribution, so
   $$\mathbb{P}(n \text{ in 3 min}) = e^{-15}\frac{15^n}{n!}, \quad n = 0, 1, 2, \ldots$$

- Numerator: Since non-overlapping intervals are independent,

$$\begin{aligned}\mathbb{P}(x \text{ in first 2min and } n \text{ in 3min}) &= \mathbb{P}(x \text{ in first 2min and } n-x \text{ in last min})\\ &= \mathbb{P}(x \text{ in first 2min}) \cdot \mathbb{P}(n-x \text{ in last min})\\ &= e^{-10}\frac{10^x}{x!} \cdot e^{-5}\frac{5^{n-x}}{(n-x)!}\end{aligned}$$

Combining and simplifying gives

$$\mathbb{P}(x \text{ in first 2min} \mid n \text{ in 3min}) = \binom{n}{x}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{n-x}, \quad x = 0, 1, \ldots, n,$$

which is the PMF of $\mathrm{Bin}(n, 2/3)$.

Problem 2.8. A continuous random variable X is said to have a Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ if it has PDF

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Use the properties of the Gamma function, namely $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ for $\alpha > 1$, to obtain $\mathbb{E}[X]$ and $\mathrm{Var}(x)$.

Solution 2.8. We can solve this problem without even knowing what the $\Gamma(\alpha)$ function is. We use a trick that we have used many times, which involves rewriting the expression as an integral of the PDF which sums to 1.

Expected Value: By definition,

$$\begin{aligned}\mathbb{E}[X] = \int_0^\infty x \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha} e^{-\frac{x}{\beta}}\, dx\\ &= \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^{\alpha} e^{-\frac{x}{\beta}}\, dx}_{=1 \text{ integral of the PDF of } \Gamma(\alpha+1,\beta) \text{ r.v.}}\end{aligned}$$

where we multiplied and divided by $\beta\Gamma(\alpha+1)$ to match the normalization terms. Since using the recursive property of the Γ function, we see that

$$\mathbb{E}[X] = \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\beta\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \beta\alpha.$$

Variance: A similar computation as above implies that

$$\begin{aligned}\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx &= \int_0^\infty x^2 \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha+1} e^{-\frac{x}{\beta}}\, dx\\ &= \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha+2)\beta^{\alpha+2}} x^{\alpha+1} e^{-\frac{x}{\beta}}\, dx}_{=1 \text{ integral of the PDF of } \Gamma(\alpha+2,\beta) \text{ r.v.}}\\ &= \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)}.\end{aligned}$$

Next, using the recursive properties of the Γ function,

$$\mathbb{E}[X^2] = \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} = \frac{\beta^2(\alpha+1)\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\beta^2(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \beta^2(\alpha+1)\alpha.$$

Therefore,

$$\mathrm{Var}(x) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \beta^2(\alpha+1)\alpha - \beta^2\alpha^2 = \beta^2\alpha.$$

Remark 13. Notice that then $\alpha = 1$, then the PDF is simply the exponential distribution. In fact, the relation is even deeper and it can be shown that the sum of n independent $\mathrm{Exp}(\beta)$ distributions has a $\Gamma(n, \beta)$ distribution.

2.5 Proofs of Key Results

Problem 2.9. Compute the mean and variance of $X \sim \mathrm{Unif}(a, b)$.

Solution 2.9. We can directly compute

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx \\ &= \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}.\end{aligned}$$

To compute the variance, we have

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_a^b x^2 \frac{1}{b-a}\,dx \\ &= \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3}.\end{aligned}$$

so after some algebra,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

Problem 2.10. Let $N(t)$ be a Poisson process with rate λ occurrences per time interval. If X is the length of time until the first occurrence, show that $X \sim \mathrm{Exp}(\lambda^{-1})$.

Solution 2.10. Since $\mathbb{P}(X \le 0) = 0$, it remains to consider $x > 0$. We start by computing the CDF of X, which is the length of time until first event occurs

$$\begin{aligned}F_X(t) &= \mathbb{P}(X \le t) \\ &= \mathbb{P}(\text{time to 1st occurrence} \le t) \\ &= 1 - \mathbb{P}(\text{time to first occurrence} > t) \\ &= 1 - \mathbb{P}(\text{no occurrence between } (0,t))\end{aligned}$$

We know how to model the number of event occurrences between time $(0, t)$ since it is $\mathrm{Poi}(\lambda t)$. Let $N(t) \sim \mathrm{Poi}(\lambda t)$. Then

$$\begin{aligned}F_X(t) &= 1 - \mathbb{P}(\text{no occurrence between } (0,t)) \\ &= 1 - \mathbb{P}(N(t) = 0) \\ &= 1 - \frac{e^{-\lambda t}(\lambda t)^0}{0!} \\ &= 1 - e^{-\lambda t}.\end{aligned}$$

So we have $F_X(t) = 1 - \exp(-\lambda t)$ for $t > 0$ and $F_X(t) = 0$ otherwise. We can take the derivative with respect to t for $t > 0$, to obtain the PDF

$$f_X(t) = \frac{d}{dt} F_X(t) = \lambda \exp(-\lambda t)$$

and $f_X(t) = 0$ for $t \le 0$. We recognize this PDF as the one corresponding to $\mathrm{Exp}(\lambda^{-1})$ (with waiting time parameter $\theta = \frac{1}{\lambda}$).

Remark 14. The (actual) Poisson process $N(t)$ counts the number of occurrences up to time t. While X is the waiting time until the first occurrence. By independent increments and homogeneous property, the waiting time between occurrences has the same distribution as the waiting time until the first occurrence. This means that the exponential distribution $\mathrm{Exp}(\lambda^{-1})$ models the waiting time between each event occurrence in a Poisson distribution with rate λ.

Problem 2.11. Let $n \ge 1$. Show that

$$\Gamma(n) = \int_0^\infty y^{n-1} e^{-y} dy = (n-1)!.$$

Solution 2.11. This follows from repeated integration by parts. Let $n \ge 1$, we have

$$\Gamma(n) = \int_0^\infty y^{n-1} e^{-y} dy.$$

Integrating by parts we see that

$$\Gamma(n) = \int_0^\infty y^{n-1} e^{-y} dy = -y^{n-1} e^{-y} \Big|_0^\infty + (n-1) \int_0^\infty y^{n-2} e^{-y} dy = (n-1)\Gamma(n-1).$$

Repeating this inductively implies that

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \cdots = (n-1)(n-2)\cdots 3 \cdot 2 \cdot \Gamma(1) = n!$$

since $\Gamma(1) = 1$ (using the fact that the integral of the PDF of $X \sim \mathrm{Exp}(1)$ is 1).

Remark 15. Notice that the integration by parts computation holds for $\alpha > 1$ which are not necessarily integer valued to conclude that $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$. The problem is that the inductive step might not lead to a nice number since $\Gamma(x)$ for $x < 1$ doesn't necessarily simplify.

Problem 2.12. Compute the mean and variance of $X \sim \mathrm{Exp}(\theta)$.

Solution 2.12. We use the change of variable $x = y\theta$ with $dx = \theta dy$

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx \overset{(x=y\theta)}{=} \int_0^\infty y e^{-y} \theta dy \\ &= \theta \underbrace{\int_0^\infty y e^{-y} dy}_{=\Gamma(2)} = \theta\Gamma(2) = \theta \cdot (1!) = \theta. \end{aligned}$$

To compute the variance, we have

$$\begin{aligned}\mathbb{E}[X^2] &= \int_0^\infty x^2 \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx \overset{(x=y\theta)}{=} \int_0^\infty \theta y^2 e^{-y} \theta dy \\ &= \theta^2 \underbrace{\int_0^\infty y^{3-1} e^{-y} dy}_{=\Gamma(3)} = \theta^2 \Gamma(3) = \theta \cdot (2!) = 2\theta^2.\end{aligned}$$

so

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 2\theta^2 - \theta^2 = \theta^2$$

Alternative Solution: We can also integrate by parts directly without using the Gamma function. Starting with

$$\mathbb{E}[X] = \int_0^\infty x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

we have by integrating by parts

$$\mathbb{E}[X] = \int_0^\infty x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx = \frac{x}{\theta} \cdot (-\theta e^{-\frac{x}{\theta}}) - \frac{1}{\theta} \theta^2 e^{-\frac{x}{\theta}} \Big|_0^\infty = \theta.$$

Next, to compute

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

we have by integrating by parts

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx = \frac{x^2}{\theta} \cdot (-\theta e^{-\frac{x}{\theta}}) - \frac{2x}{\theta} \cdot \theta^2 e^{-\frac{x}{\theta}} + \frac{2}{\theta} \cdot (-\theta^3 e^{-\frac{x}{\theta}}) \Big|_0^\infty = 2\theta^2.$$

Problem 2.13. Prove the memoryless property: Theorem 2.

Solution 2.13. Recall the CDF of $X \sim \mathrm{Exp}(\theta)$ is

$$F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^x f_X(t) dt = \int_0^x \theta^{-1} e^{-t/\theta}\, dt = 1 - e^{-x/\theta}, \quad x > 0.$$

Therefore,

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \le x) = 1 - F_X(x) = e^{-x/\theta}.$$

Hence,

$$\begin{aligned}\mathbb{P}(X > s+t \,|\, X > s) &= \frac{\mathbb{P}(X > s+t \text{ and } X > s)}{\mathbb{P}(X > s)} \\ &= \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > s)} = \frac{e^{-(s+t)/\theta}}{e^{-s/\theta}} \\ &= e^{-t/\theta} = \mathbb{P}(X > t)\end{aligned}$$

as desired.

Problem 2.14. Prove that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Solution 2.14. We have

$$\Gamma\Big(\frac{1}{2}\Big) = \int_0^\infty y^{\frac{1}{2}-1}e^{-y}dy = \int_0^\infty y^{-\frac{1}{2}}e^{-y}dy$$

We can do the change of variables $y = \frac{x^2}{2}$, $dy = x\,dx$ so

$$\Gamma\Big(\frac{1}{2}\Big) = \sqrt{2}\int_0^\infty e^{-\frac{x^2}{2}}\,dx.$$

By Fubini's theorem and a change of variables into polar coordinates

$$\begin{aligned}\Gamma\Big(\frac{1}{2}\Big)^2 = 2\bigg(\int_0^\infty e^{-\frac{x^2}{2}}\,dx\bigg)^2 = 2\int_0^\infty e^{-\frac{x^2}{2}}\,dx\int_0^\infty e^{-\frac{y^2}{2}}\,dy &= 2\int_0^\infty\int_0^\infty e^{-\frac{x^2+y^2}{2}}\,dxdy\\ &= 2\int_0^{\frac{\pi}{2}} e^{-\frac{r^2}{2}}r\,drd\theta\\ &= 2\int_0^{\frac{\pi}{2}} d\theta = \pi.\end{aligned}$$

So $\Gamma\Big(\frac{1}{2}\Big) = \sqrt{\pi}$.

Remark 16. A modification of this argument can be used to compute the normalization constant of a standard Gaussian random variable.

Problem 2.15. Prove Theorem 3.

Solution 2.15. We have

$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \le z\right) = \mathbb{P}(X \le \sigma z + \mu).$$

By the chain rule, for all $z \in \mathbb{R}$,

$$F_Z'(z) = \frac{d}{dz}F_X(\sigma z + \mu) = \sigma f_X(\sigma z + \mu) = \frac{\sigma}{\sqrt{2\pi\sigma^2}}e^{\frac{-(\sigma z+\mu-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} = \varphi(x)$$

which we recognize as the PDF of the standard normal distribution. In particular, we have shown that $Z = \frac{X-\mu}{\sigma}$ has standard normal distribution.

Problem 2.16. If $X \sim \mathrm{N}(\mu, \sigma^2)$ show that

$$\mathbb{E}[X] = \mu \qquad \mathrm{Var}(X) = \sigma^2.$$

Solution 2.16. We first do the computations when $Z \sim \mathrm{N}(0,1)$. We have

$$\mathbb{E}[Z] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty xe^{\frac{-x^2}{2}} = 0$$

because $xe^{\frac{-x^2}{2}}$ is an odd function. Next, to compute the second moment, we can integrate by parts

$$\mathbb{E}[Z^2] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty x^2e^{\frac{-x^2}{2}} = -\frac{x}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\bigg|_{-\infty}^\infty + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{\frac{-x^2}{2}}\,dx = 1$$

where the boundary term vanishes because $e^{\frac{-x^2}{2}}$ goes to zero faster than x and the second term is the integral of the PDF of a standard normal. Hence we have shown that

$$\mathbb{E}[Z] = 0 \qquad \operatorname{Var}(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = 1.$$

To get the mean and variance of $X \sim \mathrm{N}(\mu, \sigma^2)$, we can use the standardization trick to write $X \sim \sigma Z + \mu$. Therefore,

$$\mathbb{E}[X] = \mathbb{E}[\sigma Z + \mu] = \sigma\,\mathbb{E}[Z] + \mu = \mu$$

and

$$\operatorname{Var}(X) = \operatorname{Var}(\sigma Z + \mu) = \sigma^2 \operatorname{Var}(Z) = \sigma^2.$$