# Chapter 6. Continuous Random Variables

To motivate continuous random variables, consider an experiment that measures $X$, how long a battery lasts before it dies.

We can ask for the probability that the battery survives for a time in any interval we want.

For example,

$\mathrm{P}(10 \text{ hr} \leq X \leq 15 \text{ hr}) = 0.45$

$\mathrm{P}(10 \text{ hr} \leq X \leq 10 \text{ hr + 1 sec}) = 0.0001$

$\mathrm{P}(10 \text{ hr} \leq X \leq 10 \text{ hr + 1 nanosec}) = 0.0000000000000001$

The idea is that as the size of the interval gets very small, the probability that $X$ has a value inside that tiny interval also becomes small. More than that, it makes sense that the probability will become proportional to the size of the interval as the interval gets smaller:

$$\begin{aligned} \text{As } \epsilon \to 0, \quad \mathrm{P}(x \leq X \leq x+\epsilon) &\propto \epsilon \\ &= f_X(x)\epsilon. \end{aligned}$$

This motivates the existence of a function $f_X(x)$ which we call the **probability density function** or **pdf**. Note that we made key use of the fact that $X$ has a continuous range.

Now we give the mathematical definition of continuous RV. Suppose again we have a random experiment with sample space $S$ and probability measure P.

Recall our definition of a random variable as a mapping $X : S \to \mathbb{R}$ from the sample space $S$ to the real numbers inducing a probability measure $\mathrm{P}_X(B) = \mathrm{P}\left(X^{-1}(B)\right), B \subseteq \mathbb{R}$.

**Definition 6.0.1.** *A random variable $X$ is (absolutely)* **continuous** *if $\exists f_X : \mathbb{R} \to \mathbb{R}$ (measurable) such that $f$ is non-negative and*

$$\mathrm{P}(X \in B) = \int_{x \in B} f_X(x)dx, \qquad B \subseteq \mathbb{R},$$

*in which case $f_X$ is referred to as the* **probability density function**, *or* **pdf**, *of $X$.*

This definition is entirely consistent with our motivation above. Just consider the set $B = [a, b]$ as $b$ approaches $a$.

$$\mathrm{P}(a \leq X \leq b) = \int_a^b f_X(x)dx,$$

$$\mathrm{P}(a \leq X \leq a+\epsilon) = \int_a^{a+\epsilon} f_X(x)dx \approx f_X(a)\epsilon, \quad \text{as } \epsilon \to 0.$$

In words,

The probability that $X$ is between $x$ and $x + dx$ is $f_X(x)dx$.

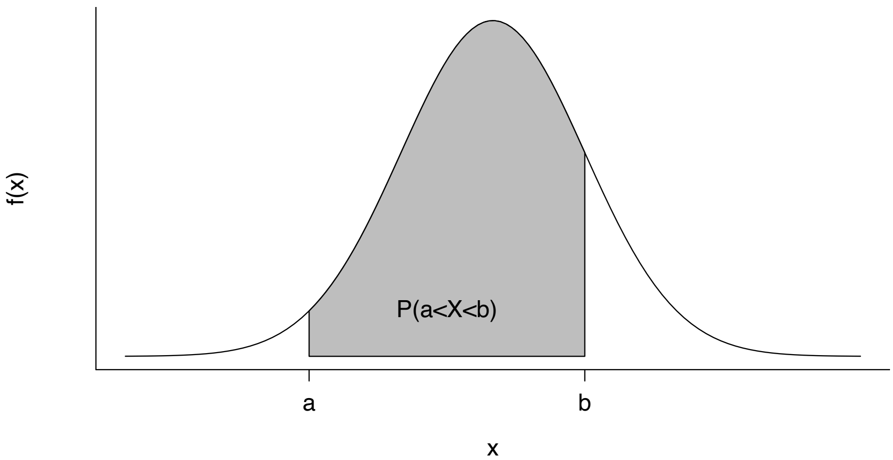
***Note*** The pdf $f_X(x)$ is NOT a probability. It is a probability *density*, having units of $1/$[units of $X$]. In the battery example, the pdf has units of $1/$seconds.

For some small interval $\epsilon$ (which has units of $X$), $f_X(x)\epsilon$ is a probability.

Visualization: Suppose we are interested in whether a continuous random variable $X$ lies in an interval $(a, b)$. From the definition of the pdf we have,

$$\mathrm{P}(a < X < b) = \int_a^b f_X(x)dx.$$

That is, the area under the pdf between $a$ and $b$.



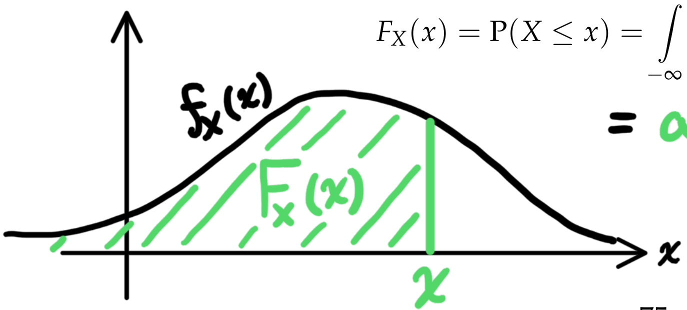

### 6.0.1 Continuous Cumulative Distribution Function

**Definition 6.0.2.** *The* **cumulative distribution function** *of* **CDF**, $F_X$ *of a continuous random variable $X$ is defined as*

$$F_X(x) = \mathrm{P}(X \le x), \quad x \in \mathbb{R}.$$

***Note*** From now on, when we speak of a continuous random variable, we will implicitly assume the absolutely continuous case. Then we can write the CDF as an integral over the pdf,

$$F_X(x) = \mathrm{P}(X \le x) = \int_{-\infty}^{x} f_X(x')dx', \quad x \in \mathbb{R}.$$




= area under $f_x$ to the left of $x$.

### 6.0.2 Properties of Continuous $F_X$ and $f_X$

By analogy with the discrete case, let $\mathbb{X}$ be the range of $X$, so that $\mathbb{X} = \{x : f_X(x) > 0\}$.

i) For the cdf of a continuous random variable,

$$\lim_{x\to-\infty} F_X(x) = 0, \quad \lim_{x\to\infty} F_X(x) = 1.$$

ii) At values of $x$ where $F_X$ is differentiable

$$f_X(x) = \left.\frac{d}{dt}F_X(t)\right|_{t=x} \equiv F'_X(x).$$

iii) If $X$ is a continuous RV,

$$\mathrm{P}(X = a) = \int_a^a f_X(x)dx = 0,$$

**Warning!** People usually forget, that $\mathrm{P}(X = x) = 0$ for all $x$ when $X$ is a continuous random variable.

iv) For $a < b$,

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

v) The pdf $f_X(x)$ is not itself a probability. Therefore, unlike the pmf of a discrete random variable, we do not require $f_X(x) \le 1$ (in fact, it doesn't even make sense to compare two quantities with different units).

vi) From Definition 6.0.1 it is clear that the pdf of a continuous random variable $X$ completely characterises its distribution, so we often just specify $f_X$.

It follows that a function $f_X$ is a pdf for a continuous random variable $X$ if and only if

i) $f_X(x) \ge 0$,

ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$ ("normalization" condition)

**Example** Consider an experiment to measure the length of time that an electrical component functions before failure. The sample space of outcomes of the experiment, $S$ is $\mathbb{R}^+$ and if $A_x$ is the event that the component functions longer than $x$ time units, suppose that $\mathrm{P}(A_x) = \exp\{-x^2\}$ for any $x > 0$.

Define continuous random variable $X : S \to \mathbb{R}^+$, by $X(s) = x \iff$ component fails at time $x$. Then, if $x > 0$

$$F_X(x) = \mathrm{P}(X \le x) = 1 - \mathrm{P}(A_x) = 1 - \exp\{-x^2\}$$

and $F_X(x) = 0$ if $x \leq 0$. Hence if $x > 0$,

$$f_X(x) = \frac{d}{dt} F_X(t) \bigg|_{t=x} = 2x \exp\{-x^2\}$$

and zero otherwise.

Figure 6.1 displays the probability density function (left) and cumulative distribution function (right). Note that both the PDF and CDF are defined for all real values of $x$, and that both are continuous functions.

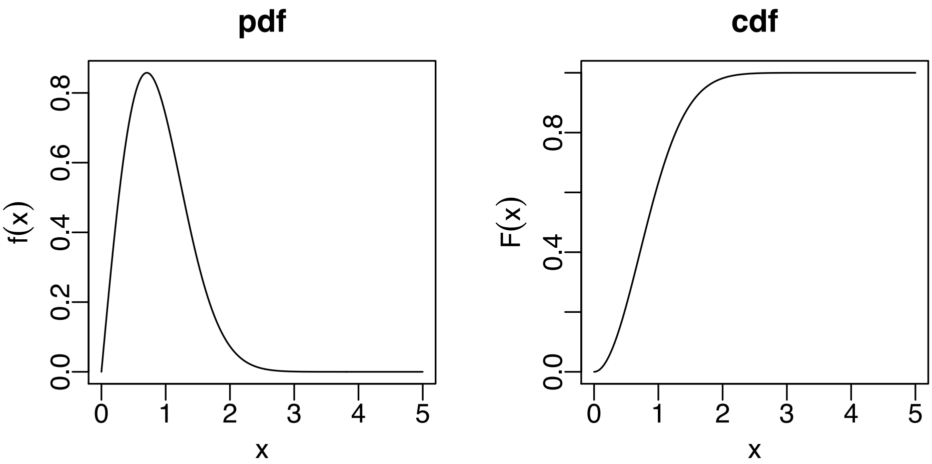



Figure 6.1: PDF $f_X(x) = 2x\exp\{-x^2\}$, $x > 0$, and CDF $F_X(x) = 1 - \exp\{-x^2\}$.

Also note that here

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{0}^{x} f_X(t)dt$$

as $f_X(x) = 0$ for $x \leq 0$, and also that

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{0}^{\infty} f_X(x)dx = 1$$

■

**Example** Suppose we have a continuous random variable $X$ with probability density function given by

$$f_X(x) = \begin{cases} cx^2, & 0 < x < 3 \\ 0, & \text{otherwise} \end{cases}$$

for some unknown constant $c$.

*Questions*

Q1) Determine $c$.

Q2) Find the cdf of $X$.

Q3) Calculate $\mathrm{P}(1 < X < 2)$.

*Solutions*

S1) We must have

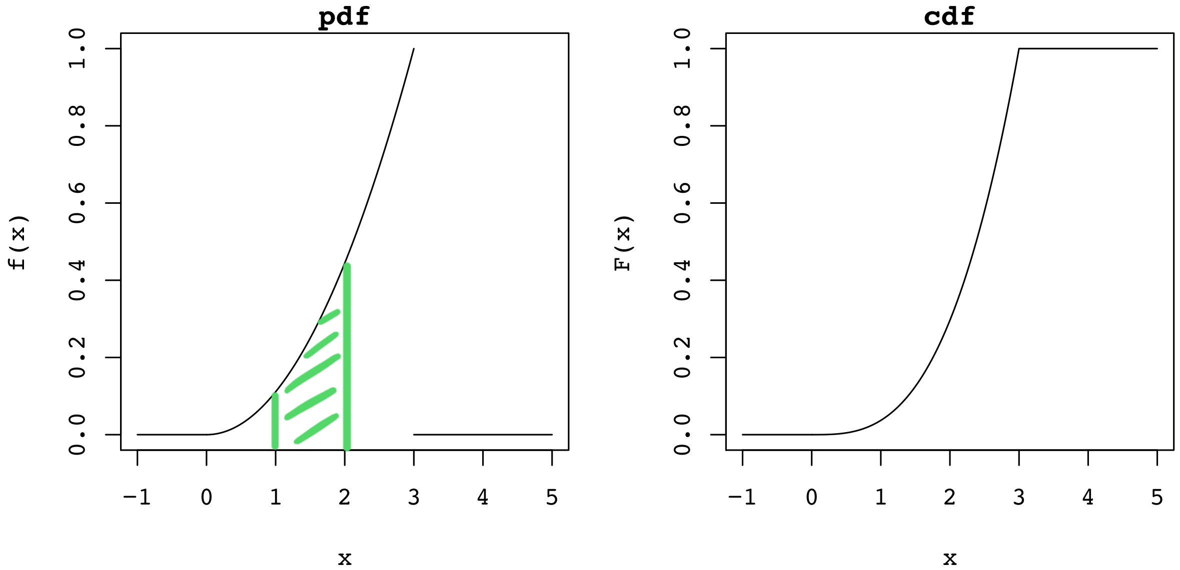
$$1 = \int_0^3 cx^2 dx = c\left[\frac{x^3}{3}\right]_0^3 = 9c$$
$$\implies c = \frac{1}{9}.$$

S2)

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \int_{-\infty}^{x} f(u)du = \int_0^x \frac{u^2}{9}du = \frac{x^3}{27} & 0 \le x \le 3 \\ 1 & x > 3 \end{cases}$$

S3)

$$\underline{\mathrm{P}(1 < \mathrm{X} < 2)} = F(2) - F(1) = \frac{8}{27} - \frac{1}{27} = \frac{7}{27} = 0.2593.$$



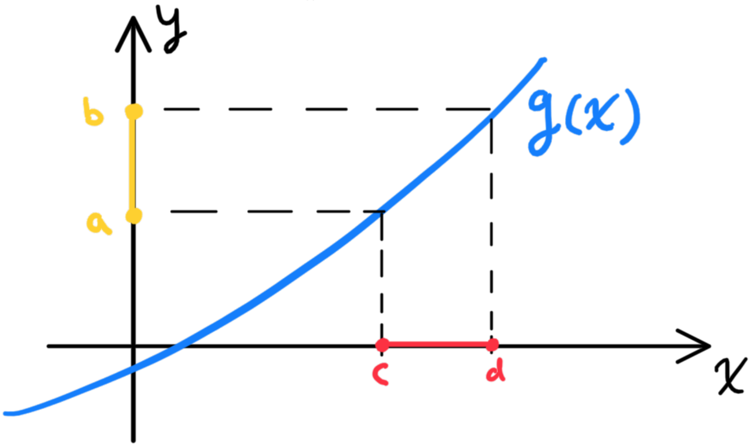

■

### 6.0.3 Transformations

Suppose that $X$ is a continuous random variable $X$ with pdf $f_X$ and cdf $F_X$. Let $Y = g(X)$ be a function of $X$ for some (measurable) function $g : \mathbb{R} \to \mathbb{R}$ s.t. $g$ is continuous. We call $Y = g(X)$ a transformation of $X$.

When the random variable $X$ takes the value $x$, the random variable $Y$ takes the value $y = g(x)$.

How can we find the pdf $f_Y$ and cdf $F_Y$ of the random variable $Y$?




$P(a<Y<b) = P(c<X<d)$

We can compute the pdf and cdf of $Y = g(X)$ by the following methods.

**Method 1:**

Step 1. Find cdf $F_Y(y)$ in terms of $F_X(x)$.
Step 2. Differentiate $F_Y(y)$ to get pdf $f_Y(y)$.

Suppose $g$ is strictly monotonically increasing (so $g^{-1}$ exists). The cdf of $Y$ is given by

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(g(X) \leq y) = \mathrm{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

$\{s \in S : g(X(s)) \leq y\}$ $\quad$ $\{s \in S : X(s) \leq g^{-1}(y)\}$

The pdf of $Y$ is given by using the chain rule of differentiation:

$$f_Y(y) = F_Y'(y) = f_X\left(g^{-1}(y)\right) g^{-1\prime}(y)$$

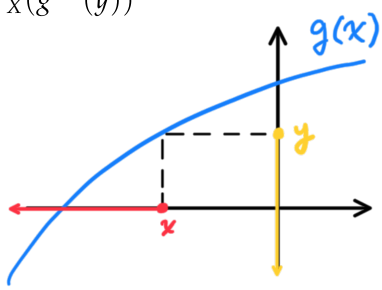
Note $g^{-1\prime}(y) = \dfrac{d}{dy} g^{-1}(y)$ is positive since we assumed $g$ was increasing.

If $g$ monotonic decreasing, we have that

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(g(X) \leq y) = \mathrm{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

By comparison with before, we would have

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



with $g^{-1'}(y)$ always negative.

Therefore, for $Y = g(X)$ we have

$$f_Y(y) = f_X\left(g^{-1}(y)\right)|g^{-1'}(y)|. \tag{6.1}$$

**Example** Let $f_X(x) = e^{-x}$ for $x > 0$. Hence, $F_X(x) = \int_0^x f_X(u)du = 1 - e^{-x}$. Let $Y = g(X) = \log(X)$. Then the range of $Y$ is $\mathbb{R}$ and

$$\begin{aligned} F_Y(y) = \mathrm{P}(Y \leq y) &= \mathrm{P}\left(\log(X) \leq y\right) \\ &= \mathrm{P}\left(X \leq \mathrm{e}^y\right) = F_X\left(\mathrm{e}^y\right). \end{aligned}$$

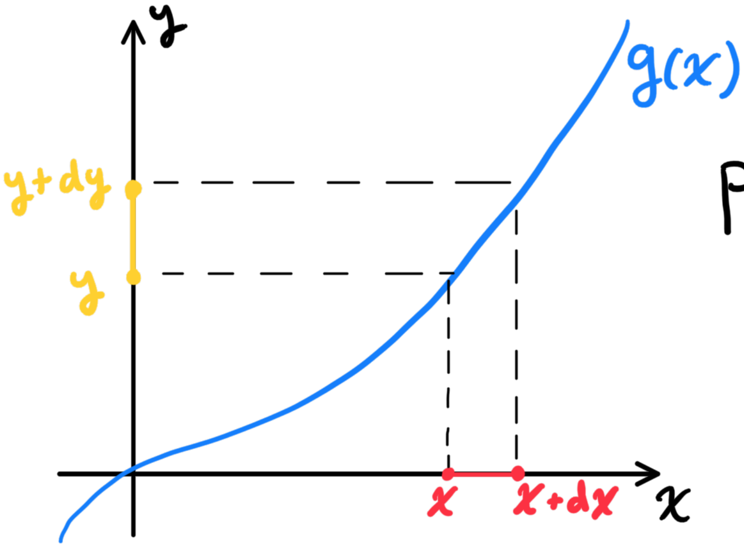
Taking the derivative of the cdf gives the pdf,

$$\begin{aligned} f_Y(y) = \frac{d}{dy}F_Y(y) &= \frac{d}{dy}F_X\left(\mathrm{e}^y\right) \\ &= F_X'\left(\mathrm{e}^y\right)\frac{d\mathrm{e}^y}{dy} \\ &= f_X\left(\mathrm{e}^y\right)\mathrm{e}^y \\ &= \mathrm{e}^{-\mathrm{e}^y}\mathrm{e}^y. \end{aligned}$$

■

**Method 2:**

Go to the pdf directly.




$$P(\underline{y<Y<y+dy}) = P(\underline{x<X<x+dx})$$

$$\boxed{f_Y(y)\,dy = f_X(x)\,dx}$$

Key: the mapping $g$ determines the relation between $y, y+dy$ and $x, x+dx$:

$$y = g(x) \qquad\qquad x = g^{-1}(y)$$
$$y + dy = g(x + dx) \qquad\qquad dy = g'(x)dx$$

**Example** Let $f_X(x) = e^{-x}$ for $x > 0$ and let $Y = g(X) = \log(X)$. Then the range of $Y$ is $\mathbb{R}$. For any $y \in \mathbb{R}$ we have $y = \log(x)$ for some $x > 0$.

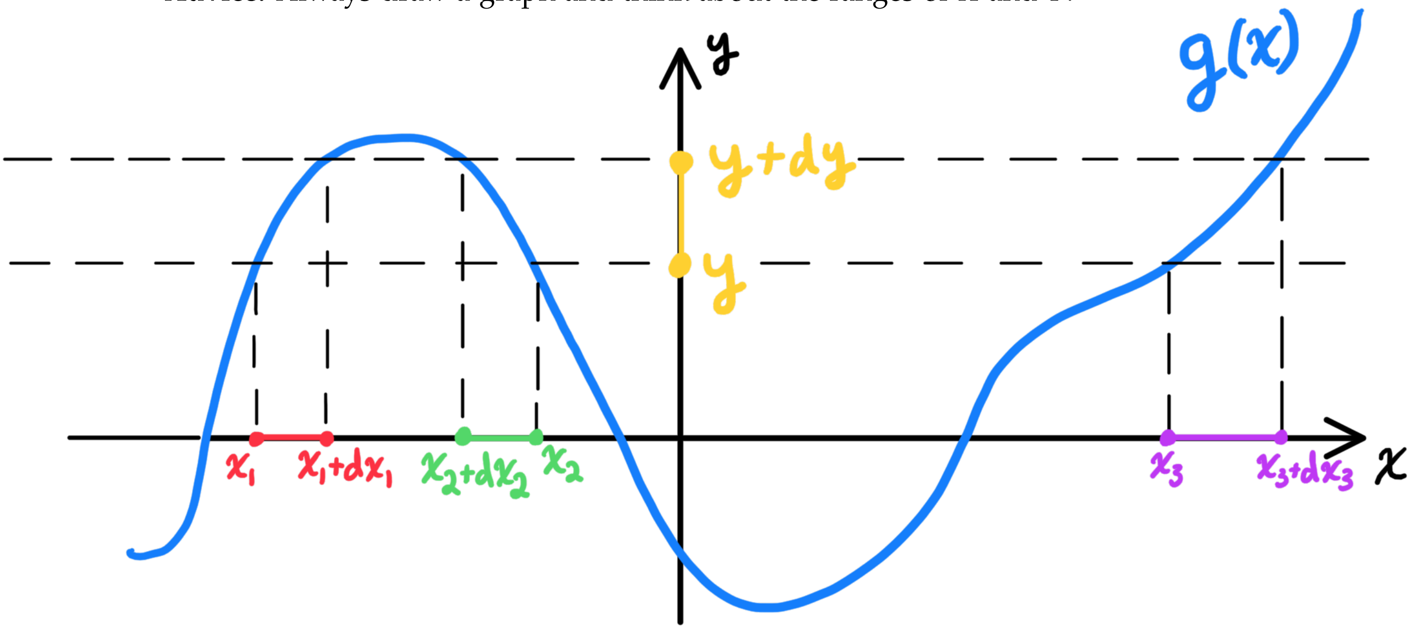
We start with the mnenomic $f_X(x)dx = f_Y(y)dy$ (remembering to always take absolute values of the differentials).

The transformation $g(\cdot)$ tells us how $dy$ and $dx$ are related: $dy = dx/x$. Or $dx = xdy = e^y dy$. Plugging into our mnemonic equation,

$$f_Y(y)dy = f_X(x)dx = e^{-x}e^y dy = e^{-e^y}e^y dy.$$

Cancelling the $dy$ from both sides gives us the pdf of $Y$. ■

**Warning:** $g(\cdot)$ may not be a 1-to-1 function, e.g. $Y = X^2$.
**Advice:** Always draw a graph and think about the ranges of $X$ and $Y$.




$$\underset{\displaystyle f_Y(y)dy}{\overset{\|}{P\begin{pmatrix}Y \text{ between } y\\ \underline{\text{and } y+dy}\end{pmatrix}}} = \underset{\displaystyle f_X(x_1)dx_1}{\overset{\|}{P\begin{pmatrix}X \text{ between } x_1\\ \underline{\text{and } x_1+dx_1}\end{pmatrix}}} + \underset{\displaystyle f_X(x_2)\underbrace{|dx_2|}_{\text{width of all intervals must be positive}}}{\overset{\|}{P\begin{pmatrix}X \text{ between } x_2\\ \underline{\text{and } x_2+dx_2}\end{pmatrix}}} + \underset{\displaystyle f_X(x_3)dx_3}{\overset{\|}{P\begin{pmatrix}X \text{ between } x_3\\ \underline{\text{and } x_3+dx_3}\end{pmatrix}}}$$

The transformation $g(\cdot)$ lets us get the $x_i$'s and $dx_i$'s in terms of $y$ and $dy$:

$$\begin{aligned} y &= g(x_1) & y &= g(x_2) & y &= g(x_3) \\ dy &= g'(x_1)dx_1 & dy &= g'(x_2)dx_2 & dy &= g'(x_3)dx_3 \end{aligned}$$

## 6.1 Mean, Variance and Quantiles

### 6.1.1 Expectation

For a discrete random variable $X$ the expectation is $\mathrm{E}(X) = \sum_{x \in \mathbb{X}} x\mathrm{P}(X = x)$, i.e. the sum, over all possible values the RV can take, of each possible value times the probability that the RV equals that value. The analogous notion for continuous random variables is

**Definition 6.1.1.** *For a continuous random variable X we define the* **mean** *or* **expectation** *of X,*

$$\mu_X \text{ or } \mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

**Extension:** More generally, for a (measurable) *function of interest* of the random variable $g : \mathbb{R} \to \mathbb{R}$ we have

$$\mathrm{E}\left(g(X)\right) = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

**Example** Let $X$ be a continuous RV with pdf

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

And define a new RV $Y$ according to $Y = g(X)$, where $g(x) = e^x$. Let's find the expected value of $Y$ in two ways.

1. First find the pdf of $Y$ and then calculate the expectation $\mathrm{E}(Y) = \int y\, f_Y(y) dy$.

   Since the range of $X$ is $[0, 1]$, the range of $Y$ is $[1, e]$. For $y \notin [1, e]$ clearly $f_Y(y) = 0$. For $y \in [1, e]$, we write

   $$f_Y(y)dy = f_X(x)dx, \text{ where } y = e^x \text{ and so } dy = e^x dx.$$

   which yields

   $$f_Y(y) = f_X(x)\frac{dx}{dy} = 1\frac{1}{e^x} = \frac{1}{y}.$$

   We use this to calculate $\mathrm{E}(Y)$.

   $$\mathrm{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{1}^{e} y\frac{1}{y} dy = \int_{1}^{e} dy = e - 1.$$

2. Alternatively we can go directly to the expectation:

   $$\mathrm{E}(Y) = \mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{0}^{1} e^x\, 1\, dx = e^x \Big|_0^1 = e - 1.$$

■

As in the discrete case, expectation is linear:

$$\mathrm{E}\left[g(X)+h(X)\right]=\mathrm{E}[g(X)]+\mathrm{E}[h(X)],$$

and

$$\mathrm{E}(aX+b)=a\mathrm{E}(X)+b,\quad (a,b\in\mathbb{R}).$$

### 6.1.2 Variance

**Definition 6.1.2.** *The* **variance** *of a continuous random variable $X$ is given by*

$$\sigma_X^2 \text{ or } \mathrm{Var}_X(X)=E[(X-\mu_X)^2]=\int_{-\infty}^{\infty}(x-\mu_X)^2 f_X(x)dx.$$

and again it is easy to show that

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

For a linear transformation $aX+b$ we again have

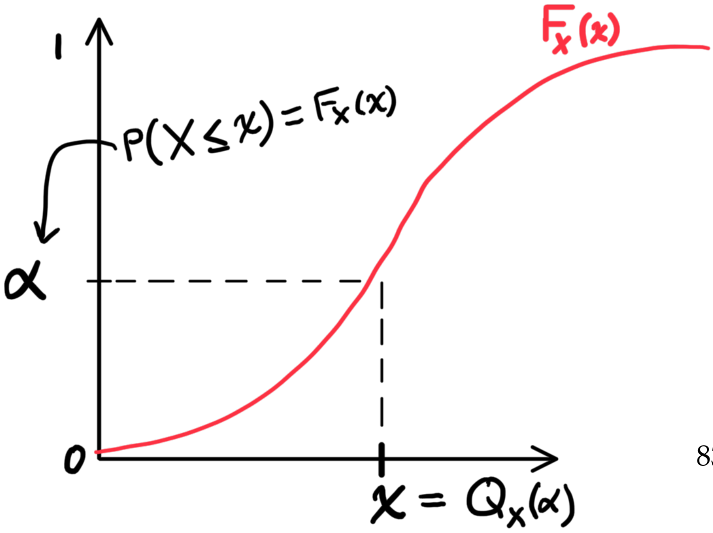
$$\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X),\qquad \forall a,b\in\mathbb{R}.$$

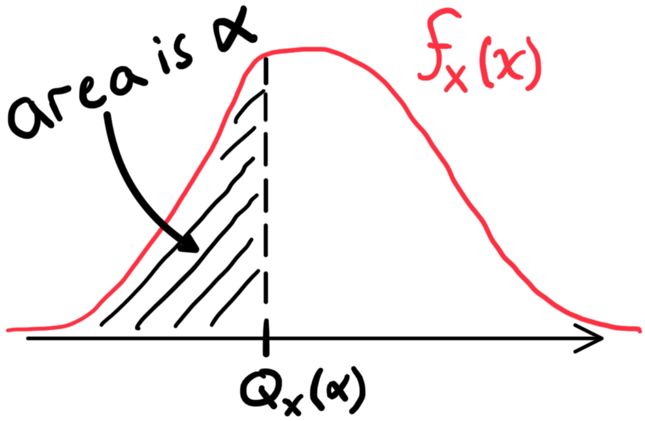
### 6.1.3 Quantiles

Recall we defined the lower and upper quartiles and median of a sample of data as points (1/4, 3/4, 1/2)-way through the ordered sample. This idea can be generalized to continuous random variables as follows:

**Definition 6.1.3.** *For a (continuous) random variable $X$ we define the* **$\alpha$-quantile** $Q_X(\alpha)$, $0\le\alpha\le 1$ *to satisfy* $\mathrm{P}(X\le Q_X(\alpha))=\alpha$,

$$Q_X(\alpha)=F_X^{-1}(\alpha).$$



In particular the **median** of a random variable $X$ is $F_X^{-1}\left(\frac{1}{2}\right)$. That is, the solution to the equation $F_X(x) = \frac{1}{2}$.

**Example** Suppose we have a continuous random variable $X$ with probability density function given by

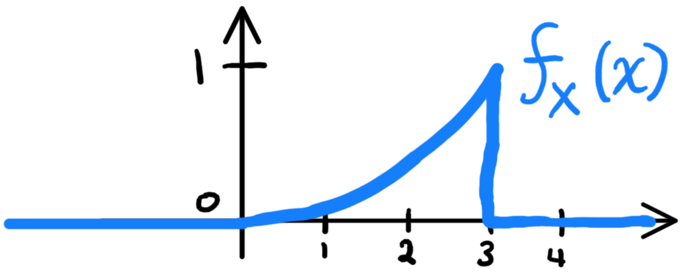
$$f_X(x) = \begin{cases} x^2/9, & 0 < x < 3 \\ 0, & \text{otherwise.} \end{cases}$$

*Questions*

Q1) Calculate $\mathrm{E}(X)$.

Q2) Calculate $\mathrm{Var}(X)$.

Q3) Calculate the median of $X$.




*Solutions*

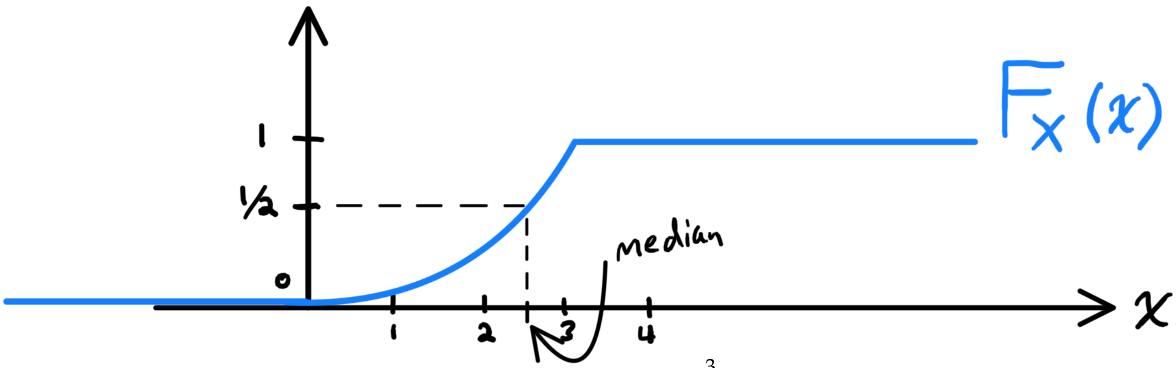
S1)

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^3 x.\frac{x^2}{9} dx = \left.\frac{x^4}{36}\right|_0^3 = \frac{3^4}{36} = 2.25$$

S2)

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x) = \int_0^3 x^2.\frac{x^2}{9} = \left.\frac{x^5}{45}\right|_0^3 = \frac{3^5}{45} = 5.4$$

So $\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 = 5.4 - 2.25^2 = 0.3375$




S3) From our earlier example, $F(x) = \frac{x^3}{27}$, for $0 < x < 3$.

Setting $F(x) = \frac{1}{2}$ and solving, we get

$$\frac{x^3}{27} = \frac{1}{2} \iff x = \sqrt[3]{\frac{27}{2}} = \frac{3}{\sqrt[3]{2}} \approx 2.3811$$

for the median.

■

## 6.2 Some Important Continuous Random Variables

### 6.2.1 Continuous Uniform Distribution

Suppose $X$ is a continuous random variable with probability density function

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise,} \end{cases}$$

Then $X$ is said to follow a uniform distribution on the interval $(a, b)$ and we write $X \sim \mathrm{U}(a, b)$.

***Notes***

- The cdf is

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

- The case $a = 0$ and $b = 1$ is referred to as the **Standard uniform**.

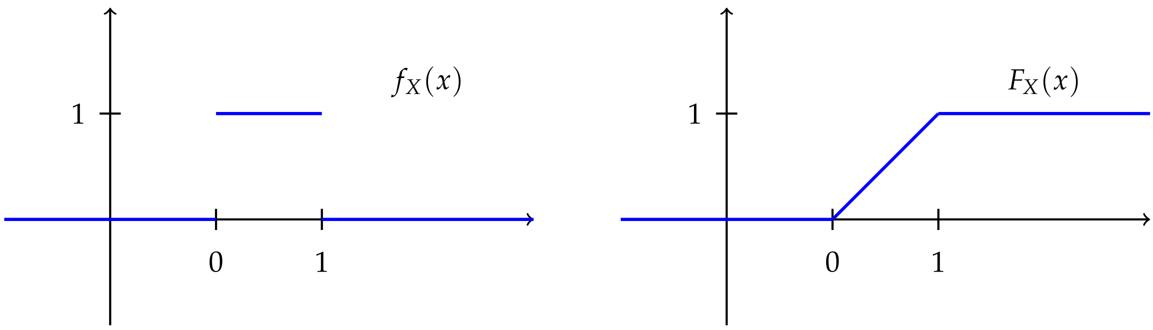



Figure 6.2: PDF and CDF of a standard uniform distribution.

- We can start with a standard uniform RV and transform it to get any uniform RV we want: Start with $X \sim \mathrm{U}(0, 1)$ and let $Y$ be given by the transformation $Y = a + (b-a)X$ (with $b > a$). Then $Y \sim \mathrm{U}(a, b)$.

  *Proof:*

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(a + (b-a)X \leq y) = \mathrm{P}\left(X \leq \frac{y-a}{b-a}\right) = F_X\left(\frac{y-a}{b-a}\right)$$

  This is 0 when $y < a$ and 1 when $y > b$. And for $a \leq y \leq b$, $F_Y(y) = \left(\dfrac{y-a}{b-a}\right)$. Therefore,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} \frac{1}{b-a} & a \leq y \leq b \\ 0 & y < a \text{ or } y > b \end{cases}$$

- To find the mean of $X \sim \mathrm{U}(a, b)$,

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x . \frac{1}{b-a} dx = \left[ \frac{x^2}{2(b-a)} \right]_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}.$$

Similarly we get $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \frac{(b-a)^2}{12}$, so

$$\mu = \frac{a+b}{2}, \qquad \sigma^2 = \frac{(b-a)^2}{12}.$$

### 6.2.2 Exponential Distribution

Suppose now $X$ is a random variable taking value on $\mathbb{R}^+ = [0, \infty)$ with pdf

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \geq 0,$$

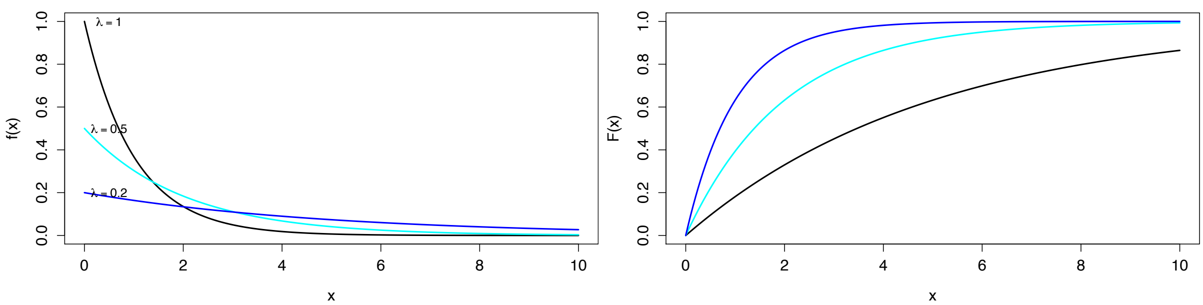
for some $\lambda > 0$.

Then $X$ is said to follow an exponential distribution with *rate* parameter $\lambda$ and we write $X \sim \mathrm{Exp}(\lambda)$.

***Notes***

- The cdf is

$$F_X(x) = \mathrm{P}(X \leq x) = 1 - e^{-\lambda x}, \quad x > 0.$$




PDFs and CDFs for Exponential distribution with different rate parameters.

- An alternative representation uses $\theta = 1/\lambda$ as the parameter of the distribution. This is sometimes used because the expectation and variance of the Exponential distributions are

$$\mathrm{E}(X) = \frac{1}{\lambda} = \theta, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2} = \theta^2.$$

- *Interpretation:* For $T \sim \text{Exp}(\lambda)$, $T$ can be interpreted as the time until an event occurs, where events occur at an "average rate" $\lambda$.

  The exponential distribution is the continuous version of the geometric distribution.

  Divide up the interval $[0, t]$ into $N$ intervals of size $dt = t/N$ (we will let $N \to \infty$). The probability of an event occurring during any segment is $\lambda dt$; that is what it means for $\lambda$ to be the "average rate" (we assume events occurring in different intervals are independent). The probability that the first event occurs at a time $\leq t$ is one minus the probability that no events occur by time $t$. I.e.

$$\mathrm{P}(T \leq t) = 1 - \lim_{N\to\infty} (1 - \lambda dt)^N = 1 - \lim_{N\to\infty} \left(1 - \frac{\lambda t}{N}\right)^N = 1 - e^{-\lambda t},$$

  which is exactly the exponential cdf.

- If we have already waited for a time $t$, what is the probability we are still waiting at time $t + s$?

$$\mathrm{P}(X > s+t|X > t) = \frac{\mathrm{P}(X > s + t \cap X > t)}{\mathrm{P}(X > t)} = \frac{\mathrm{P}(X > s + t)}{\mathrm{P}(X > t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathrm{P}(X > s).$$

  Thus, for all $x, t > 0$, $\mathrm{P}(X > t + x|X > t) = \mathrm{P}(X > x)$ — this is known as the **Lack of Memory Property**, and is unique to the exponential distribution amongst continuous distributions.

  *Interpretation:* If we think of the exponential variable as the time to an event, then knowledge that we have waited time $s$ for the event tells us nothing about how much longer we will have to wait – the process has *no memory*.

- Exponential random variables are often used to model the time until occurrence of a random event where there is an assumed constant risk ($\lambda$) of the event happening over time, and so are frequently used as a simplest model, for example, in reliability analysis. So examples include:

  – the time to failure of a component in a system;

  – the time until a radioactive particle decays;

  – the distance we travel along a road until we find the next pothole;

  – the time until the next jobs arrives at a database server;

  Notice the duality between some of the exponential random variable examples and those we saw for a Poisson distribution. In each case, "number of events" has been replaced with "time between events".

  *Claim:*

  Suppose we have a random process where $\forall x > 0$, the number of events $N_x$ occurring in an interval of size $x$ is a Poisson RV with an average rate $\lambda$, i.e. $N_x \sim \text{Poisson}(\lambda x)$

(called a *homogeneous Poisson process*). Then the separation between two events is an exponential RV with parameter $\lambda$.

*Proof:* An event occurs at some $x_1$. The probability that the next event does not occur in the next interval of length $x$ is $\mathrm{P}(N_x = 0) = e^{-\lambda x}(\lambda x)^0/0! = e^{-\lambda x}$. I.e. the probability that the next event occurs at a time less than $x$ is $1 - e^{-\lambda x}$, exactly the exponential cdf.

### 6.2.3 Normal (Gaussian) Distribution

Suppose $X$ is a random variable taking values in $\mathbb{R}$ with pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

for some $\mu \in \mathbb{R}, \sigma^2 > 0$. Then $X$ is said to follow a Gaussian or normal distribution with mean $\mu$ and variance $\sigma^2$, and we write $X \sim \mathrm{N}(\mu, \sigma^2)$.

***Notes***

- The cdf of $X \sim \mathrm{N}(\mu, \sigma^2)$ is not analytically tractable for any $(\mu, \sigma)$, so we can only write

$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt.$$

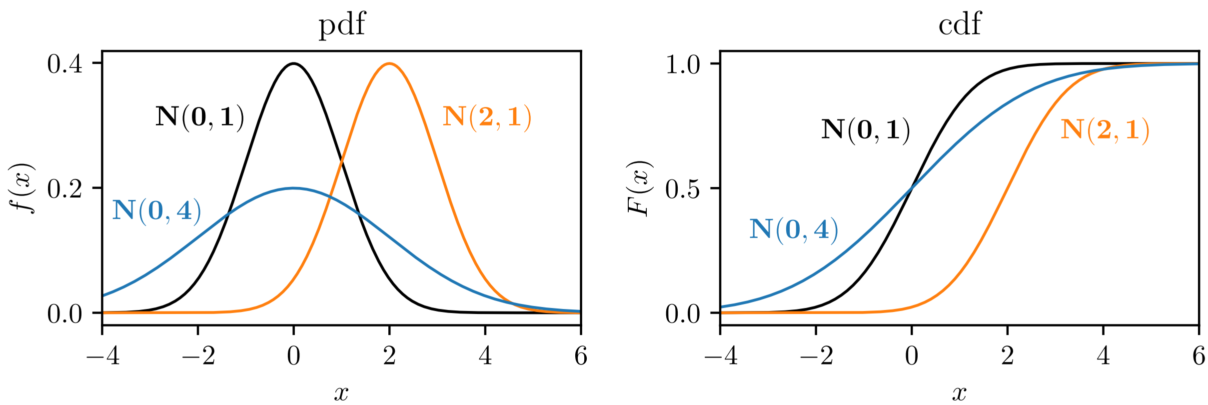



Figure 6.3: PDFs and CDFs of normal distributions with different means and variances.

- Special Case: If $\mu = 0$ and $\sigma^2 = 1$, then $X$ has a **standard normal distribution** (sometimes called the **unit normal distribution**). The pdf of the standard normal distribution is written as $\phi(x)$ and simplifies to

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}.$$

The cdf of the standard normal distribution is written as $\Phi(x)$. Again, the integral cannot be solved analytically and we can only write

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

- If $X \sim \mathrm{N}(0,1)$, and

$$Y = \sigma X + \mu$$

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

Re-expressing this result: if $Y \sim N(\mu, \sigma^2)$ and $X = (Y - \mu)/\sigma$, then $X \sim \mathrm{N}(0,1)$. This is an important result as it allows us to write the cdf of any normal distribution in terms of $\Phi$:

If $Y \sim \mathrm{N}(\mu, \sigma^2)$ and we set $X = \dfrac{Y - \mu}{\sigma}$, $F_X(x) = \Phi(x)$ and we can write the cdf of an arbitrary normal random variable $Y$ in terms of the standard normal cdf:

$$\begin{aligned} F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}\left(\sigma X + \mu \leq y\right) = \mathrm{P}\left(X \leq \frac{y - \mu}{\sigma}\right) \\ = \Phi\left(\frac{y - \mu}{\sigma}\right). \end{aligned}$$

- Since the cdf, and therefore any probabilities, associated with a normal distribution are not analytically available, numerical integration procedures are used to find approximate probabilities. In particular, statistical tables contain values of the standard normal cdf $\Phi(z)$ for a range of values $z \in \mathbb{R}$, and the quantiles $\Phi^{-1}(\alpha)$ for a range of values $\alpha \in (0,1)$. Linear interpolation is used for approximation between the tabulated values. As seen in the point above, all normal distribution probabilities can be related back to probabilities from a standard normal distribution.

- Table 6.1 is an example of a statistical table for the standard normal distribution.

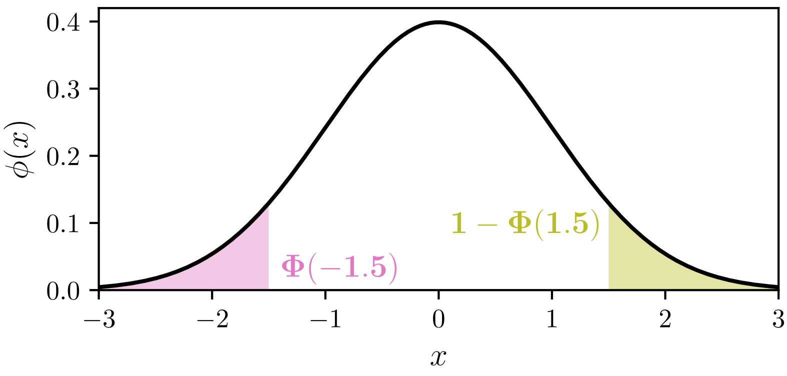
| $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ |
|---|---|---|---|---|---|---|---|
| 0 | 0.5 | 0.9 | 0.816 | 1.8 | 0.964 | 2.8 | 0.997 |
| 0.1 | 0.540 | 1.0 | 0.841 | 1.9 | 0.971 | 3.0 | 0.998 |
| 0.2 | 0.579 | 1.1 | 0.864 | 2.0 | 0.977 | 3.5 | 0.9998 |
| 0.3 | 0.618 | 1.2 | 0.885 | 2.1 | 0.982 | 1.282 | 0.9 |
| 0.4 | 0.655 | 1.3 | 0.903 | 2.2 | 0.986 | 1.645 | 0.95 |
| 0.5 | 0.691 | 1.4 | 0.919 | 2.3 | 0.989 | 1.96 | 0.975 |
| 0.6 | 0.726 | 1.5 | 0.933 | 2.4 | 0.992 | 2.326 | 0.99 |
| 0.7 | 0.758 | 1.6 | 0.945 | 2.5 | 0.994 | 2.576 | 0.995 |
| 0.8 | 0.788 | 1.7 | 0.955 | 2.6 | 0.995 | 3.09 | 0.999 |

Table 6.1

Notice that $\Phi(z)$ has been tabulated for $z > 0$.

This is because the standard normal pdf $\phi$ is *symmetric* about 0, that is, $\phi(-z) = \phi(z)$. For the cdf $\Phi$, this means

$$\Phi(z) = 1 - \Phi(-z).$$




So, for example, $\Phi(-1.2) = 1 - \Phi(1.2) \approx 1 - 0.885 = 0.115$.

Similarly, if $Z \sim \mathrm{N}(0,1)$ and we want, for example, $\mathrm{P}(Z > 1.5) = 1 - \mathrm{P}(Z \le 1.5) = 1 - \Phi(1.5)(= \Phi(-1.5))$.

So more generally we have

$$\mathrm{P}(Z > z) = \Phi(-z).$$

We will often have cause to use the 97.5% and 99.5% quantiles of $\mathrm{N}(0,1)$, given by $\Phi^{-1}(0.975)$ and $\Phi^{-1}(0.995)$.

$$\Phi(1.96) \approx 97.5\%.$$

So with 95% probability an $\mathrm{N}(0,1)$ random variable will lie in $[-1.96, 1.96]$ ($\approx [-2,2]$).

$$\Phi(2.58) = 99.5\%.$$

So with 99% probability an $\mathrm{N}(0,1)$ random variable will lie in $[-2.58, 2.58]$.

**Example** An analogue signal received at a detector (measured in microvolts) may be modelled as a Gaussian random variable $X \sim \mathrm{N}(200, 256)$.

*Questions*

Q1) What is the probability that the signal will exceed $240\mu V$?

Q2) What is the probability that the signal is larger than $240\mu V$ given that it is greater than $210\mu V$?

*Solutions*

S1)

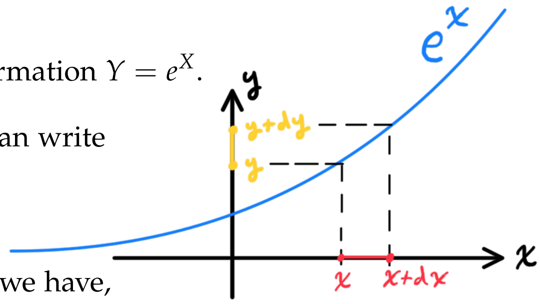
$$\mathrm{P}(X > 240) = 1 - \mathrm{P}(X \le 240) = 1 - \Phi\left(\frac{240 - 200}{\sqrt{256}}\right) = 1 - \Phi(2.5) \approx 0.00621.$$

S2)

$$\mathrm{P}(X > 240 | X > 210) = \frac{\mathrm{P}(X > 240)}{\mathrm{P}(X > 210)} = \frac{1 - \Phi\left(\frac{240-200}{\sqrt{256}}\right)}{1 - \Phi\left(\frac{210-200}{\sqrt{256}}\right)} \approx 0.02335.$$

■

**Example** Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$, and consider the transformation $Y = e^X$.




This is a 1-to-1 mapping from $\mathbb{X} = \mathbb{R}$ to $\mathbb{Y} = \mathbb{R}^+$ and we can write

$$\underline{f_Y(y)dy} = \underline{f_X(x)dx},$$

where $y = e^x$. Then $dy = e^x dx$, or $dx = dy/e^x = dy/y$ and we have,

$$f_Y(y) = \frac{f_X(\log y)}{y} = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log y - \mu)^2}{2\sigma^2}\right], \quad y > 0,$$

and we say $Y$ follows a **log-normal** distribution. Note that the mean and variance of $Y$ are *not* equal to $\mu$ and $\sigma^2$. ■

## Sums of many random variables

Let $X_1, X_2, \ldots, X_n$ be $n$ **independent and identically distributed (i.i.d.)** random variables from **any** probability distribution, each with mean $\mu$ and variance $\sigma^2$.

We will show in the next chapter that

$$\mathrm{E}\left(\sum_{i=1}^n X_i\right) = n\mu, \quad \mathrm{Var}\left(\sum_{i=1}^n X_i\right) = n\sigma^2.$$

First notice

$$\mathrm{E}\left(\sum_{i=1}^n X_i - n\mu\right) = 0, \quad \mathrm{Var}\left(\sum_{i=1}^n X_i - n\mu\right) = n\sigma^2.$$

Dividing by $\sqrt{n}\sigma$, we obtain

$$\mathrm{E}\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}\right) = 0, \quad \mathrm{Var}\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}\right) = 1.$$

**Theorem 6.3** (Central Limit Theorem or CLT)**.**

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \sim \mathrm{N}(0,1).$$

Dividing the numerator and denominator on the lefthand side by $n$, this can also be written as

$$\lim_{n\to\infty} \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1), \quad \text{where} \quad \overline{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

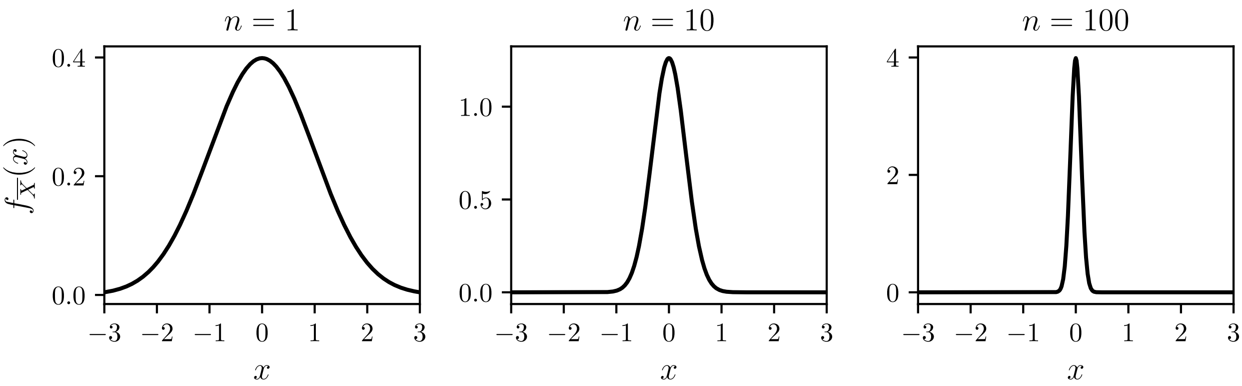
Another way of saying this is that when $n$ is large we have approximately

$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

or

$$\sum_{i=1}^n X_i \sim \mathrm{N}\left(n\mu, n\sigma^2\right).$$

We note here that although all these approximate distributional results hold irrespective of the distribution of the $\{X_i\}$, in the special case where $X_i \sim \mathrm{N}(\mu, \sigma^2)$ these distributional results are, in fact, exact. This is because the sum of independent normally distributed random variables is also normally distributed.




**Example** Consider the most simple example, that $X_1, X_2, \ldots$ are i.i.d. Bernoulli$(p)$ discrete random variables taking value 0 or 1.

Then the $\{X_i\}$ have mean $\mu = p$ and variance $\sigma^2 = p(1-p)$. By definition, we know that for any $n$ we have

$$\sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p).$$

which has mean $np$ and variance $np(1-p)$.

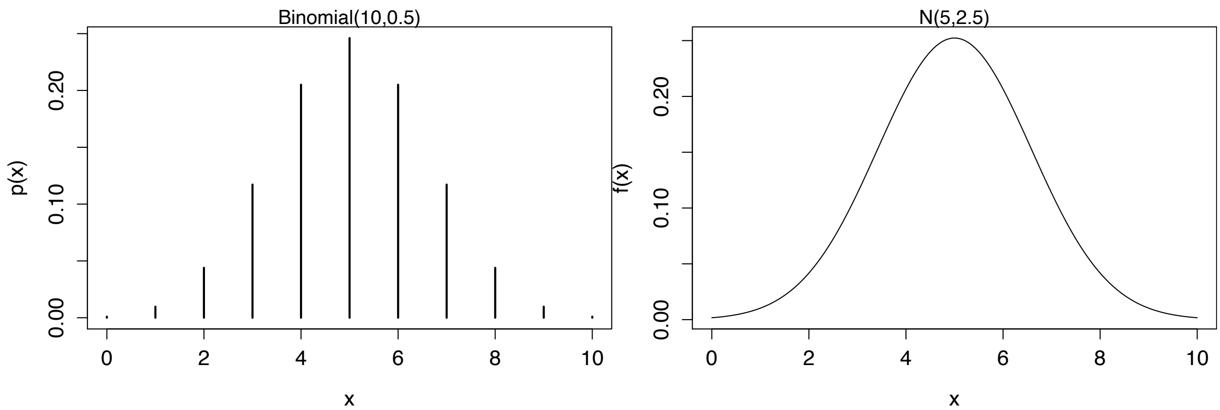
But now, by the Central Limit Theorem (CLT), we also have for large $n$ that approximately:

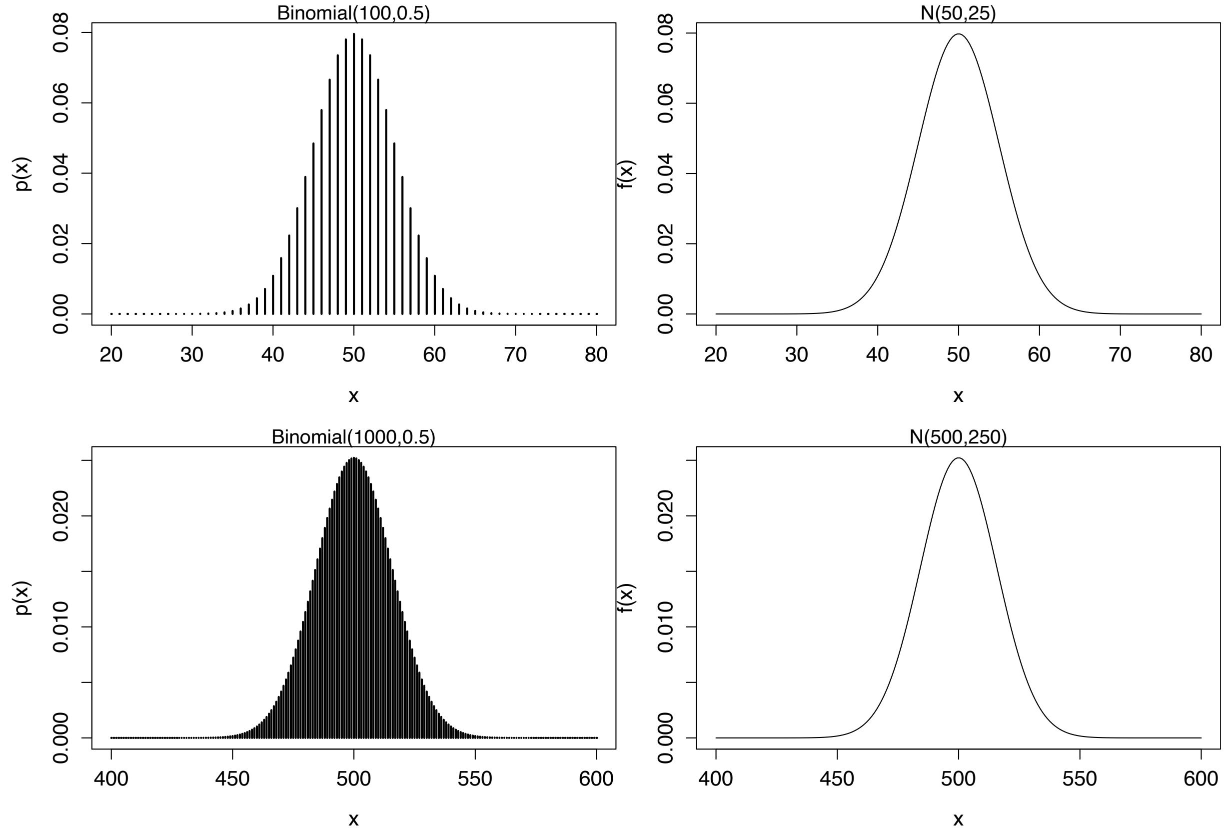
$$\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(n\mu, n\sigma^2\right) \equiv \mathrm{N}(np, np(1-p)).$$

So for large n

$$\text{Binomial}(n, p) \approx \mathrm{N}(np, np(1-p)).$$

Notice that the LHS is a discrete distribution, and the RHS is a continuous distribution.






$\blacksquare$

**Example** Suppose $X$ was the number of heads found on 1000 tosses of a fair coin, and we were interested in $\mathrm{P}(X \leq 490)$.

Using the binomial distribution pmf, we would need to calculate

$$\mathrm{P}(X \leq 490) = p_X(0) + p_X(1) + p_X(2) + \ldots + p_X(490)(\approx 0.27).$$

However, using the CLT we have approximately $X \sim \mathrm{N}(500, 250)$ and so

$$\mathrm{P}(X \leq 490) \approx \Phi\left(\frac{490 - 500}{\sqrt{250}}\right) = \Phi(-0.632) = 1 - \Phi(0.632) \approx 0.26.$$

$\blacksquare$

## 6.4 Further Examples

**Example** $X$ is the temperature in $^\circ F$ at which a random chemical reaction occurs,

$$f_X(x) = xe^{-x^2/2}, \text{ for } x > 0.$$

Let $Y$ be the temperature in $^\circ C$, so $Y = \dfrac{5}{9}(X - 32)$.

*Question* Find $f_Y(y)$.

First figure out the range of $Y$. Since $X > 0$, $Y$ can only take values greater than $-\frac{5}{9}(32)$.

<u>Method 1:</u>

Get CDF of $Y$; differentiate with respect to $y$ to get PDF of $Y$.

$$\begin{aligned} F_Y(y) = \mathrm{P}(Y \le y) = \mathrm{P}\left(\frac{5}{9}(X - 32) \le y\right) && \text{using the transformation definition} \\ = \mathrm{P}\left(X \le \frac{9}{5}y + 32\right) && \text{rearranging the inequality} \\ = F_X\left(\frac{9}{5}y + 32\right). && \text{definition of cdf} \end{aligned}$$

Now differentiate with respect to $y$:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X\left(\frac{9}{5}y + 32\right) \\ &= f_X\left(\frac{9}{5}y + 32\right)\frac{9}{5} \\ &= \frac{9}{5}\left(\frac{9}{5}y + 32\right)e^{-\left(\frac{9}{5}y+32\right)^2/2}. \end{aligned}$$

Don't forget the range

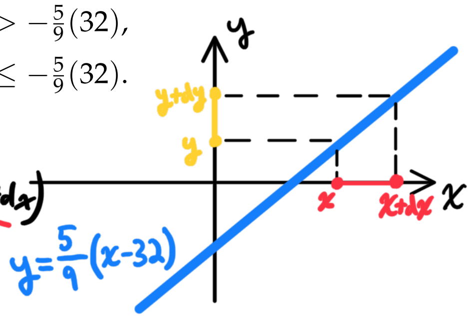
$$f_Y(y) = \begin{cases} \frac{9}{5}\left(\frac{9}{5}y + 32\right)e^{-\left(\frac{9}{5}y+32\right)^2/2} & y > -\frac{5}{9}(32), \\ 0 & y \le -\frac{5}{9}(32). \end{cases}$$

<u>Method 2:</u>

Start with

$P(y<Y<y+dy)$ $P(x<X<x+dx)$

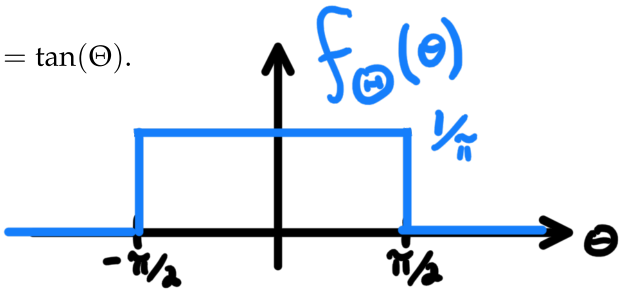
$$f_Y(y)dy = f_X(x)dx,$$




where $y = \frac{5}{9}(x - 32)$. The differential of this equation relates $dx$ to $dy$, i.e. $dy = \frac{5}{9}dx$. Write $x$ and $dx$ in terms of $y$ and $dy$, plug into the first equation, cancel the $dy$'s to find $f_Y(y)$. Don't forget the range of $Y$.

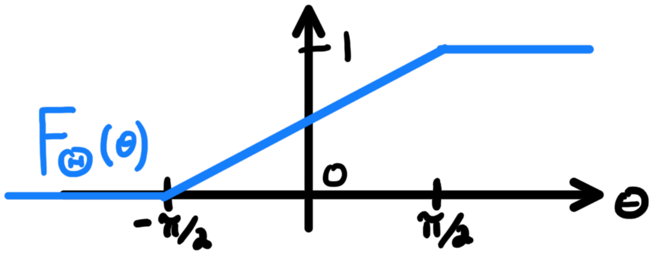
■

**Example** Suppose $\Theta \sim \text{Uniform}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, and let $X = \tan(\Theta)$.



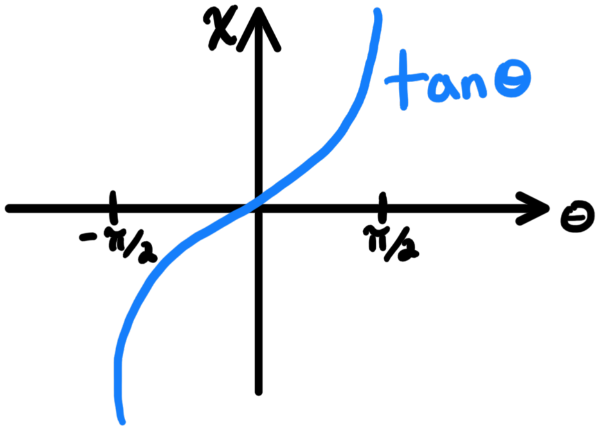

*Question* Find $F_\Theta(\theta)$.

$$F_\Theta(\theta) = \begin{cases} 0 & \theta < -\frac{\pi}{2} \\ \frac{1}{\pi}\left(\theta + \frac{\pi}{2}\right) & -\frac{\pi}{2} \le \theta \le \frac{\pi}{2} \\ 1 & \theta > \frac{\pi}{2} \end{cases}$$




*Question* Derive $F_X(x)$.

$$\begin{aligned} F_X(x) &= \mathrm{P}(X \le x) \\ &= \mathrm{P}(\tan\Theta \le x) \\ &= \mathrm{P}\left(\Theta \le \tan^{-1}(x)\right) \\ &= F_\Theta\left(\tan^{-1}(x)\right). \end{aligned}$$




*Question* Establish $f_X(x)$.

$f_X(x) = \dfrac{d}{dx} F_\Theta\left(\tan^{-1}(x)\right) = f_\Theta\left(\tan^{-1}(x)\right) \dfrac{1}{1+x^2} = \dfrac{1}{\pi(1+x^2)}$, using the derivative of inverse tangent, $\frac{d}{dx}\tan^{-1}(x) = 1/(1+x^2)$.

Don't forget the range. The transformation is continuous and goes from $x = -\infty$ when $\theta = -\pi/2$ to $x = \infty$ when $\theta = \pi/2$. Therefore the above formula for the pdf applies for all $x$. Incidentally, the distribution for $X$ is called a *Cauchy distribution*.

(Method 2 — going directly to the pdf of $X$)

Start with $f_X(x)dx = f_\Theta(\theta)d\theta$,

where $x$ and $\theta$ are related by $x = \tan\theta$. Therefore, $dx = \dfrac{1}{\cos^2\theta}d\theta$, which can be reexpressed as $dx = (1+x^2)d\theta$ using the trig identity $\tan^2\theta + 1 = 1/\cos^2\theta$. Plugging into the above,

$$f_X(x)(1+x^2)d\theta = \frac{1}{\pi}d\theta.$$

Cancelling the $d\theta$'s and rearranging gives $f_X(x)$ as above.

■

**Example** $X \sim \text{Uniform}(-1, 3)$. Let $Y = X^2$.

*Question* Find $f_Y(y)$.

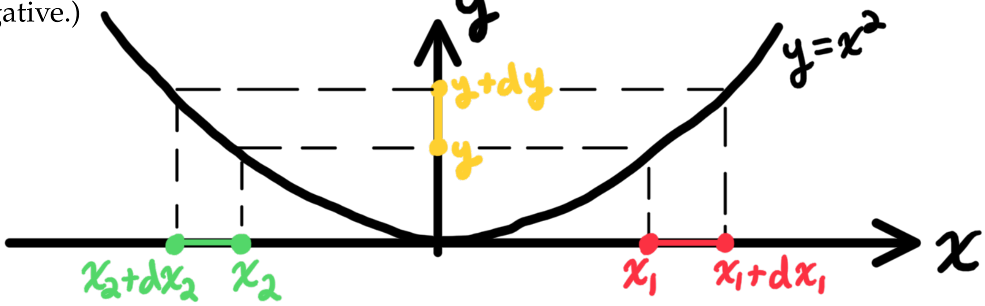
First the range of $Y$. $X$ can take any value between $-1$ and 3. Therefore, $Y$ can take any value from $Y = 0$ (when $X = 0$) to $Y = 9$ (when $X = 3$).

For each value of $Y = y$ there are potentially two values of $X$ which contribute to the probability: $x_1 = \sqrt{y}$ and $x_2 = -\sqrt{y}$.

Start with:

$$f_Y(y)dy = f_X(x_1)dx_1 + f_X(x_2)|dx_2|, \qquad (*)$$

where the $x_i$'s are related to $y$ by $x_1 = \sqrt{y}$ and $x_2 = -\sqrt{y}$. Taking differentials leads to $dx_1 = \dfrac{dy}{2\sqrt{y}}$ and $dx_2 = -\dfrac{dy}{2\sqrt{y}}$. (Note the $|dx_2|$ above, because when $y$ increases, $x_2$ decreases, so $dx_2$ is negative.)




For $0 \le y \le 1$:

For $y$'s between 0 and 1, both $x_1 = \sqrt{y}$ and $x_2 = -\sqrt{y}$ are in the range of $X$ (i.e. $-1 \le x_1, x_2 \le 3$). Therefore, since $X$ is a Uniform RV, $f_X(x_1) = f_X(x_2) = 1/4$. Plugging into equation (*) above,

$$f_Y(y)dy = \frac{1}{4}\frac{dy}{2\sqrt{y}} + \frac{1}{4}\left|\frac{-dy}{2\sqrt{y}}\right| = \frac{1}{4\sqrt{y}}dy.$$

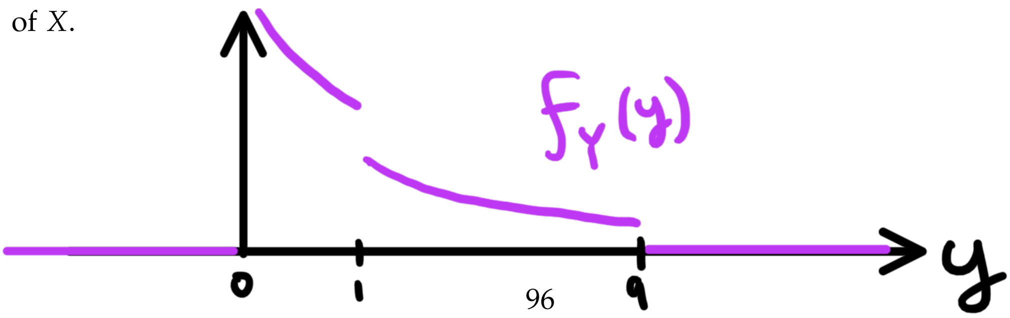
Cancelling the $dy$'s gives $f_Y(y) = 1/(4\sqrt{y})$ for $0 \le y \le 1$.

For $1 \le y \le 9$:

For these $y$'s, $x_1 = \sqrt{y}$ is in the range of $X$ but $x_2 = -\sqrt{y}$ is not. The pdf of $X$ at $x_2$ is therefore zero. So $f_X(x_1) = 1/4$ but $f_X(x_2) = 0$. Plugging into (*),

$$f_Y(y)dy = \frac{1}{4}\frac{dy}{2\sqrt{y}} + 0\left|\frac{-dy}{2\sqrt{y}}\right| = \frac{1}{8\sqrt{y}}dy.$$

Cancelling the $dy$'s gives $f_Y(y) = 1/(8\sqrt{y})$ for $1 < y \le 9$. For $y$'s less than 0 or greater than 9, the pdf of $f_Y$ is zero because both $x_1$ and $x_2$ are outside the range of $X$.




■