Lecture 2 – Probability Recap

Last Modified: August 27, 2026

Outline:

2.1 Random Variables

- Each observation X we will collect is a random variable.
- The random sample $\underline{X} = (X_1, \ldots, X_n)$ is a vector of independent random variables
- Each statistic $T(\underline{X})$, e.g. $T(\underline{X}) = \sum_{i=1}^n X_i$, is a function of random variables so it is a random variable.

Let X be a random variable describing a phenomenon of interest. The realizations x of X are real numbers. i.e., $x \in \mathbb{R}$.

Definition 1. The **distribution function** of a random variable x is a function F_X such that

$$F_X(x) = \mathbb{P}(X \leq x), \quad \text{for all } x \in \mathbb{R}$$

It is also called the **cumulative distribution function** (cdf).

Example 1. The **Exponential distribution** with parameter $\lambda > 0$ has cdf

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases} \tag{2.1}$$

We can write the cdf more compactly using indicator function notation. We write indicators in the following two equivalent ways:

$$\mathbb{1}_{\{x \in A\}} = \mathbb{1}_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases}$$

Then the cdf in (2.1) can be expressed compactly as

$$F_X(x) = \left(1 - e^{-\lambda x}\right) \mathbb{1}_{\{x \geq 0\}}$$

- If $F_X(x)$ is a continuous function then we say that X is a continuous random variable.
- If $F_X(x)$ is a step function then we say that X is discrete.

2.1.1 Continuous Random Variables

- The **probability density function** (pdf) of a continuous random variable X is defined as

$$f_X(x) = \frac{d}{dx} F_X(x), \quad \forall x \in \mathbb{R}$$

- Knowing the pdf is equivalent to knowing the distribution:

$$F_X(x) = \int_{-\infty}^{x} f_X(y) dy, \qquad \text{and} \qquad \mathbb{P}(X \in [a,b]) = \int_a^b f_X(x)\, dx$$

- Basic properties of a pdf:

$$f(x) \geq 0 \qquad \text{and} \qquad \int_{-\infty}^{+\infty} f(x) dx = 1$$

- The **Normal distribution** with mean μ and variance σ^2, denoted $\mathcal{N}(\mu, \sigma^2)$ has density

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

  See here for more on the normal.

- The **Gamma distribution** with shape α and rate β, denoted $\mathsf{Gamma}(\alpha, \beta)$ has density

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\{x \geq 0\}}$$

  where $\alpha > 0$, $\beta > 0$, and $\Gamma(\alpha)$ is the Gamma function:

  - $\Gamma(1) = 1$, $\Gamma(0) = +\infty$
  - $\Gamma(x+1) = x\Gamma(x)$ for all $x \in \mathbb{R}$
  - $\Gamma(x) = (x-1)! = 1 \cdot 2 \cdot 3 \cdots (x-1)$ for all $x \in \mathbb{N}$

  See here for more on the gamma distribution.

- The **uniform distribution** on interval (a, b) is denoted by $\mathsf{Unif}(a, b)$ and has pdf

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{\{x \in (a,b)\}}$$

  See here for more on the uniform distribution.

- The **Beta distribution** with parameters α and β is denoted by $\mathsf{Beta}(\alpha, \beta)$ and has pdf

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{1}_{\{x \in (0,1)\}}$$

  If $\alpha = \beta = 1$ then the $\mathsf{Beta}(1, 1)$ distribution has pdf

$$f_X(x) = \mathbb{1}_{\{x \in (0,1)\}}$$

  which is $\mathsf{Unif}(0, 1)$. See here for more on the beta distribution.

Example 2. Let $\alpha = 1$. Let us compute the distribution function of $X \sim \mathsf{Gamma}(1, \beta)$. By definition, $F_X(x) = \mathbb{P}(X \le x)$.

If $x \le 0$ then

$$\mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(y)\,dy = \int_{-\infty}^{x} \beta e^{-\beta y} \mathbb{1}_{\{y \ge 0\}}\,dy = \int_{-\infty}^{x} \beta 0\,dy = 0.$$

If $x \ge 0$ then

$$\begin{aligned}\mathbb{P}(X \le x) &= \int_{-\infty}^{x} f_X(y)\,dy = \int_{-\infty}^{x} \beta e^{-\beta y} \mathbb{1}_{\{y \ge 0\}}\,dy \\ &= \int_{0}^{x} \beta e^{-\beta y}\,dy = -e^{-\beta y}\Big|_0^x = 1 - e^{-\beta x}.\end{aligned}$$

Thus

$$F_X(x) = \left(1 - e^{\beta x}\right) \mathbb{1}_{\{x \ge 0\}}$$

and so $\mathsf{Gamma}(1, \beta)$ is the exponential distribution with rate parameter β.

2.1.2 Discrete Random Variables

- The **probability mass function** (pmf) of a discrete random variable X is

$$f_X(x) = \mathbb{P}(X = x), \quad x = 0, 1, 2, \ldots$$

- Knowing the pmf is equivalent to knowing the distribution:

$$F_X(x) = \sum_{y \le x} f_X(y), \qquad \text{and} \qquad \mathbb{P}(X \in [a, b]) = \sum_{x\,:\,x \in [a,b]} f_X(x)$$

- Basic properties of a pmf

$$f_X(x) \ge 0 \qquad \text{and} \qquad \sum_{x=0}^{\infty} f_X(x) = 1$$

- The **Bernoulli** distribution with parameter p denoted by $\mathsf{Ber}(p)$ has pmf

$$f_X(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

- The **Binomial** distribution with parameters n and p denoted by $\mathsf{Bin}(n, p)$ has pmf

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n$$

  See here for more on the binomial.

- The **Poisson** distribution with parameter λ denoted by $\mathsf{Poisson}(\lambda)$ has pmf

$$f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

  See here for more on the Poisson.

2.1.3 Moments and Expectation

- The **mean** of a random variable X is

$$\mathbb{E}[X] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x f_X(x)\,dx, & \text{continuous case} \\ \displaystyle\sum_{x=0}^{\infty} x f_X(x), & \text{discrete case} \end{cases}$$

  It is also called the **expected value** or **first moment**

- The k-th moment of X is given by

$$\mathbb{E}\left[X^k\right] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x^k f_X(x)\,dx, & \text{continuous case} \\ \displaystyle\sum_{x=0}^{\infty} x^k f_X(x), & \text{discrete case} \end{cases}$$

- Let g be any function. The **expected value** of $g(X)$ is given by

$$\mathbb{E}[g(X)] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} g(x) f_X(x)\,dx, & \text{continuous case} \\ \displaystyle\sum_{x=0}^{\infty} g(x) f_X(x), & \text{discrete case} \end{cases}$$

Example 3. Let $X \sim \mathsf{Gamma}(\alpha, \beta)$. The expected value is

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^{\infty} x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\,dx \\ &= \int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha} e^{-\beta x}\,dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha} e^{-\beta x}\,dx. \end{aligned}$$

To solve this integral, we recognize that $x^\alpha e^{-\beta x}$ is the *kernel* of the density of a Gamma distribution with parameters $\alpha+1$ and β, i.e.,

$$\begin{aligned} \int_0^{\infty} x^{\alpha} e^{-\beta x}\,dx &= \int_0^{\infty} x^{(\alpha+1)-1} e^{-\beta x}\,dx \\ &= \underbrace{\frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}}}_{=1} \int_0^{\infty} x^{(\alpha+1)-1} e^{-\beta x}\,dx \\ &= \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \underbrace{\int_0^{\infty} \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\beta x}\,dx}_{=1} = \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}}. \end{aligned}$$

Thus,

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

See here for more on the "massage and squint" approach of recognizing a kernel.

Example 4. $X \sim \mathsf{Poisson}(\lambda)$. The expectation is

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x f_X(x) = \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!}$$

For $x = 0$ the summand is zero. Thus, we can start the summation from 1, and do the following steps:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!} && \text{using } x/(x!) = 1/(x-1)! \\
&= \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} && \text{pull out one power of } \lambda \\
&= \lambda \underbrace{\sum_{y=0}^{\infty} \frac{\lambda^y e^{-\lambda}}{y!}}_{=1} && \text{reindex: } y = x - 1 \\
&= \lambda.
\end{aligned}$$

This is another example of "massage and squint"!

2.2 Multivariate Random Variables

- The random sample $\underline{X} = (X_1, \dots, X_n)$ is a vector of random variables, so we need to work with more than one variable at the same time.
- We focus below on several continuous random variables, but the same ideas extend to the discrete setting by replacing the density function with the mass function.

2.2.1 Pairs of random variables

For a pair of random variables (X, Y) the joint distribution is characterized by the joint density function

$$f_{X,Y}(x, y)$$

such that

$$\mathbb{P}(X \in [a_1, b_1], Y \in [a_2, b_2]) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{X,Y}(x, y)\, dy\, dx$$
$$\mathbb{P}(X \leq a, Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y)\, dy\, dx$$

The distributions of X and Y (called the **marginal** distributions) can be found as

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \qquad \text{and} \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx \tag{2.2}$$

Let us explain why $f_X(x)$ has to be as in (2.2). On the the one hand,

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y \leq +\infty) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dy \right) dx$$

On the other hand, we also know that

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

So matching the integrands recovers the formula: $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dy$. The same reasoning applies for $f_Y(y)$.

Example 5. Consider the joint density

$$f(x,y) = \begin{cases} 6xy^2, & 0 < x < 1 \text{ and } 0 < y < 1 \\ 0, & \text{else} \end{cases}$$

What is the marginal distributions of X and Y?

For $0 < x < 1$,

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \int_0^1 6xy^2\, dy = 6x \int_0^1 y^2\, dy = 6x \left.\frac{y^3}{3}\right|_0^1 = 2x$$

Thus $f_X(x) = 2x\mathbb{1}_{\{0<x<1\}}$.

For $0 < y < 1$,

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \int_0^1 6xy^2\, dx = 6y^2 \int_0^1 x\, dx = 6y^2 \left.\frac{x^2}{2}\right|_0^1 = 3y^2$$

Thus $f_Y(y) = 3y^2\mathbb{1}_{\{0<y<1\}}$.

- Often we observe one of the variables (e..g, $Y = y$) and want to the distribution of X conditional on $Y = y$. We define the **conditional density function** as

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

- The **conditional expectation** becomes

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{+\infty} x f_{X|Y}(x \mid y)\, dx$$

  Note that this is a function of y.

Example 6. Consider the joint density

$$f(x,y) = e^{-y}\mathbb{1}_{\{0<x<y<+\infty\}}$$

Let us compute $\mathbb{E}[Y \mid X = x]$.

First we find the marginal distribution of X. For $x \geq 0$,

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy = \int_0^{+\infty} e^{-y}\mathbb{1}_{\{x\leq y\}}dy = \int_x^{+\infty} e^{-y}\, dy = -e^{-y}\Big|_x^{+\infty} = e^{-x}$$

We recognize that X has an exponential distribution with parameter 1. Thus,

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} \mathbb{1}_{\{x<y\}} = e^{-(y-x)} \mathbb{1}_{\{x<y\}}$$

The conditional expectation is then

$$\begin{aligned}
\mathbb{E}[Y \mid X = x] &= \int_{-\infty}^{+\infty} y f_{Y|X}(y \mid x)\, dy = \int_{x}^{+\infty} y e^{-(y-x)} dy \\
&= \int_0^{+\infty} (z+x) e^{-z} dz = \int_0^{+\infty} z e^{-z}\, dz + x \int_0^{+\infty} e^{-z}\, dz = 1 + x
\end{aligned}$$

where the third equality follows from the substitution $z = y - x$.

Definition 2. Random variables X and Y are **independent** if and only if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

2.2.2 More than two random variables

In the **multidimensional** case (more than two variables) the distribution of vector $\underline{X} = (X_1, \dots, X_n)$ of random variables is characterized by the **joint density function**

$$f_{\underline{X}}(x_1, \dots, x_n)$$

such that

$$\mathbb{P}(X_1 \in [a_1, b_1], \dots, X_n \in [a_n, b_n]) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{\underline{X}}(x_1, \dots, x_n)\, dx_1 \dots dx_n$$

For a random vector $(\underline{X}, Y) = (X_1, \dots, X_n, Y)$ with joint density $f_{\underline{X},Y}(x,y)$ the conditional density is

$$f_{\underline{X}|Y}(\underline{x} \mid y) = \frac{f_{\underline{X},Y}(x_1, \dots, x_n, y)}{f_Y(y)}$$

Definition 3. Random variables $X_1, \dots, X_n$ are **independent** if and only if

$$f_{\underline{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

2.3 Properties of Expectation

- The expected value is **linear**. This means that

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] \qquad \text{and} \qquad \mathbb{E}[cX] = c\,\mathbb{E}[X]$$

  for any random variables X and Y and any constant c. Combining these properties yields:

$$\mathbb{E}\left[\sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} c_i \mathbb{E}[X_i]$$

- If random variables X and Y are **independent** then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

- If $X_1, \dots, X_n$ are independent random variables then

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i]$$

Definition 4. The **variance** of a random variable is given by

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

This is a measure of how much X deviates from its mean $\mathbb{E}[X]$.

Lemma 2.1. We have $\mathsf{Var}(X) = \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2$

Proof.

$$\begin{aligned}
\mathsf{Var}(X) &= \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] \\
&= \mathbb{E}\Big[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\Big] \\
&= \mathbb{E}\big[X^2\big] + \mathbb{E}\Big[-2X\mathbb{E}[X]\Big] + \mathbb{E}\Big[\mathbb{E}[X]^2\Big] && \text{linearity of expectation} \\
&= \mathbb{E}\big[X^2\big] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 && -2\mathbb{E}[X] \text{ and } \mathbb{E}[X]^2 \text{ are constants} \\
&= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2
\end{aligned}$$

□

Properties of the variance:

- $\mathsf{Var}(X) \geq 0$ and $\mathsf{Var}(X) = 0$ if and only if $X = c$ for a constant c.
- $\mathsf{Var}(cX) = c^2 \,\mathsf{Var}(X)$.
- If X and Y are independent then $\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y)$.
- If $X_1, \dots, X_n$ are independent then

$$\mathsf{Var}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 \,\mathsf{Var}(X_i)$$

Example 7. Consider again the joint density

$$f(x, y) = \begin{cases} 6xy^2, & 0 < x < 1 \text{ and } 0 < y < 1 \\ 0, & \text{else} \end{cases}$$

Let us compute $\mathsf{Var}(2X - 3Y)$

- Previously, we computed the marginal densities as

$$f_X(x) = 2x\mathbb{1}_{\{0<x<1\}} \qquad \text{and} \qquad f_Y(y) = 3y^2\mathbb{1}_{\{0<y<1\}}$$

- Observing that

$$f_X(x) f_Y(y) = cxy^2 \mathbb{1}_{\{0<x<1\}} \mathbb{1}_{\{0<y<1\}}$$

  we see that X and Y are independent, and thus

$$\mathsf{Var}(2X - 3Y) = \mathsf{Var}(2X) + \mathsf{Var}(-3Y) = 4\,\mathsf{Var}(X) + 9\,\mathsf{Var}(Y)$$

- Next we compute the first and second moments:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_0^1 2x^2\,dx = 2\frac{x^3}{3}\bigg|_0^1 = \frac{2}{3} \\
\mathbb{E}\big[X^2\big] &= \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx = \int_0^1 2x^3\,dx = 2\frac{x^4}{4}\bigg|_0^1 = \frac{1}{2} \\
\mathbb{E}[Y] &= \int_{-\infty}^{+\infty} y f_Y(y)\,dy = \int_0^1 3y^3\,dy = 3\frac{y^4}{4}\bigg|_0^1 = \frac{3}{4} \\
\mathbb{E}\big[Y^2\big] &= \int_{-\infty}^{+\infty} y^2 f_Y(y)\,dx = \int_0^1 3y^4\,dy = 3\frac{y^5}{5}\bigg|_0^1 = \frac{3}{5}
\end{aligned}$$

- Putting everything back together, we have

$$\mathsf{Var}(X) = \frac{1}{2} - \frac{4}{9} = \frac{1}{18} \qquad \text{and} \qquad \mathsf{Var}(Y) = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$$

so

$$\mathsf{Var}(2X - 3Y) = \frac{4}{18} + \frac{27}{80}$$

2.4 Transformations of Random Variables

- Let X be a random variable. Often we are in a function $g(X)$, which is again a random variable.
- Let X have sample space / domain $\mathcal{X}$ (which is either $\mathbb{R}$ or $\mathbb{N}$) and let $g\colon \mathcal{X} \to \mathcal{Y}$ be a function (e.g., $g(x) = x^2$). Then $Y = g(X)$ is a random variable with domain $\mathcal{Y}$.
- How do we obtain the distribution of Y? Observe that for any set $A \subset \mathcal{Y}$,

$$\mathbb{P}(Y \in A) = \mathbb{P}(g(X) \in A) = \mathbb{P}(X \in \{x \in \mathcal{X} \mid g(x) \in A\}) = \mathbb{P}\big(X \in g^{-1}(A)\big)$$

where $g^{-1}(A) = \{x \in \mathcal{X} \mid g(x) \in \mathcal{A}\}$ is the set of values in the domain that map to points in the set A.

- If for every y there exists a single x such that $g(x) = y$, the g is **invertible** and we can define inverse $g^{-1}\colon \mathcal{Y} \to \mathcal{X}$.

Discrete case: If $\mathcal{X}$ is finite or countable then

$$\mathcal{Y} = \{y \mid y = g(x) \quad \text{for some } x \in \mathcal{X}\}$$

is finite or countable and so Y is also a discrete random variable. The probability mass function is

$$f_Y(y) = \mathbb{P}(Y = y) = \sum_{x \in g^{-1}(\{y\})} \mathbb{P}(X = x) = \sum_{x \in g^{-1}(\{y\})} f_X(x)$$

Example 8. If $X \sim \mathsf{Poisson}(\lambda)$ and $Y = 2X$, then

$$\mathcal{X} = \{0, 1, 2, \ldots,\} \quad \Longrightarrow \quad \mathcal{Y} = \{0, 2, 4, \ldots\}$$

Moreover, if $y \in \mathcal{Y}$ then

$$f_Y(y) = \sum_{x \in g^{-1}(\{y\})} f_X(x) = f_X(y/2) = \frac{\lambda^{y/2} e^{-\lambda}}{(y/2)!}$$

Here, we see that $g(x) = 2x$ is invertible with $g^{-1}(y) = y/2$.

Continuous case: a bit more complicated. Define

$$\mathcal{X} = \{x \mid f_X(x) > 0\} \quad \text{and} \quad \mathcal{Y} = \{y \mid y = g(x) \quad \text{for some } x \in \mathcal{X}\}$$

Theorem 1 (Change of variable)**.** Let g be strictly monotone (increasing or decreasing) and assume the inverse g^{-1} has a continuous derivative on $\mathcal{Y}$. Then

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|.$$

Note: Under the assumptions of the theorem, g is strictly monotone and so the inverse g^{-1} is well-defined.

Proof. First suppose that g is increasing. Then,

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

Differentiating both sides gives

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) \\ &= f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) && \text{chain rule} \\ &= f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| && g \text{ is increasing} \end{aligned}$$

Next suppose that g is decreasing. In this case,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y) = \mathbb{P}\big(X \ge g^{-1}(y)\big) \\ &= 1 - \mathbb{P}\big(X \le g^{-1}(y)\big) = 1 - F_X(g^{-1}(y)) \end{aligned}$$

and so differentiating gives

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) = \frac{d}{dy}[1 - F_X(g^{-1}(y))] \\ &= -f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) && \text{chain rule} \\ &= f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| && g \text{ is decreasing} \end{aligned}$$

$\square$

Example 9. Suppose $X \sim \mathsf{Gamma}(\alpha, \beta)$ and $Y = X^2$.

- $\mathcal{X} = \mathbb{R}_+$ and $g(x) = x^2$ is increasing on $\mathbb{R}_+$ with $g^{-1}(y) = \sqrt{y}$.
- Moreover,
$$\frac{d}{dy}g^{-1}(y) = \frac{d}{dy}\sqrt{y} = \frac{1}{2\sqrt{y}}$$
- Thus,
$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}(\sqrt{y})^{\alpha-1}e^{-\beta\sqrt{y}}\frac{1}{2\sqrt{y}}\mathbb{1}_{\{y>0\}} \\ &= \frac{\beta^\alpha}{2\Gamma(\alpha)}y^{\frac{\alpha}{2}-1}e^{-\beta\sqrt{y}}\mathbb{1}_{\{y>0\}} \end{aligned}$$

The change of variable formula is not always applicable since g may not be monotone. Another approach is to work with the cumulative distribution function (cdf) $F_Y(y)$ and then differentiate to obtain the density $f_Y(y)$.

Example 10. Let $X \sim \mathcal{N}(0,1)$. Find the distribution of $Y = X^2$.

- $\mathcal{X} = \mathbb{R}$ and $g(x) = x^2$ is *not* monotone!.
- Let us try to compute $F_Y(y)$ with $y \in \mathcal{Y} = \mathbb{R}_+$.

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}\big(X^2 \le y\big) \\ &= \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \mathbb{P}(X \le \sqrt{y}) - \mathbb{P}(X < -\sqrt{y}) && \text{since } [-\sqrt{y}, \sqrt{y}] = \{x \le \sqrt{y}\} \setminus \{x \le -\sqrt{y}\} \\ &= \mathbb{P}(X \le \sqrt{y}) - \mathbb{P}(X \le -\sqrt{y}) && X \text{ is continuous} \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

- Differentiating,

$$\begin{aligned} f_Y(y) = \frac{d}{dy} F_Y(y) &= \frac{d}{dy} F_X(\sqrt{y}) - \frac{d}{dy} F_X(-\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \end{aligned}$$

- Note that this relation holds for any continuous X!
- Recalling that $X \sim \mathcal{N}(0,1)$, we have $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and so

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} + \frac{1}{2\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{y}{2}} \end{aligned}$$

- Finally, we conclude that

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{y}{2}} \mathbb{1}_{\{y \ge 0\}}$$

  This is the Gamma distribution with parameters $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{2}$ (note that $\Gamma(1/2) = \sqrt{\pi}$).

The distribution of Y in the previous example is also called the **chi-squared distribution** with 1 degree of freedom. More generally, the **chi-squared distribution** with k-degrees of freedom corresponds to $\mathsf{Gamma}(\frac{k}{2}, \frac{1}{2})$ and has density

$$f_X(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{\{x \ge 0\}}$$

It is denoted by χ^2_k.

Example 11. Let $X_i \overset{\text{iid}}{\sim} \mathsf{Unif}(0,1)$ for $1 \le i \le n$, that is, $X_1, \ldots, X_n$ are independent and identically distributed according to $\mathsf{Unif}(0,1)$. Find the distribution of $Y = \max\{X_1, \ldots, X_n\}$.

- Notice that $\mathcal{Y} = (0,1)$. For $y \in \mathcal{Y}$,

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leq y) \\
&= \mathbb{P}(\max\{X_1, \ldots, X_n\} \leq y) \\
&= \mathbb{P}(X_1 \leq y, \ldots, X_n \leq y) \\
&= \mathbb{P}(X_1 \leq y)\mathbb{P}(X_2 \leq y) \cdots \mathbb{P}(X_n \leq y) && \text{independent} \\
&= \mathbb{P}(X_1 \leq y)^n && \text{identically distributed} \\
&= (F_{X_1}(y))^n
\end{aligned}$$

- Differentiating,

$$\begin{aligned}
f_Y(y) = \frac{d}{dy} f_Y(y) &= \frac{d}{dy}(F_{X_1}(y))^n \\
&= n(F_{X_1}(y))^{n-1} \cdot \left(\frac{d}{dy} F_{X_1}(y)\right) \\
&= n(F_{X_1}(y))^{n-1} \cdot f_{X_1}(y).
\end{aligned}$$

  Note that this relation holds for any continuous distribution of X_1.

- In our case,

$$\begin{aligned}
f_{X_1}(y) &= \mathbb{1}_{\{y \in (0,1)\}} \\
F_{X_1}(y) &= \mathbb{P}(X_1 \leq y) = \int_{-\infty}^{y} f_{X_1}(x)\, dx = \int_0^y 1\, dx = y
\end{aligned}$$

- Thus,

$$\begin{aligned}
f_Y(y) &= ny^{n-1}\mathbb{1}_{\{y \in (0,1)\}} = ny^{n-1}(1-y)^{1-1}\mathbb{1}_{\{y \in (0,1)\}} \\
&= \frac{\Gamma(n+1)}{\Gamma(1)\Gamma(n)} y^{n-1}(1-y)^{1-1}\mathbb{1}_{\{y \in (0,1)\}}
\end{aligned}$$

  and so $Y \sim \mathsf{Beta}(n, 1)$.

2.5 Properties of Special Distributions

2.5.1 Normal Distribution

- Let $X \sim \mathcal{N}(\mu, \sigma^2)$. The expectation and variance of X are

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

  The normal distribution is determined completely by its mean and variance.

- Let $X \sim \mathcal{N}(\mu, \sigma^2)$. For every constant c,
  1. $Y = c + X$ has distribution $\mathcal{N}(\mu + c, \sigma^2)$
  2. $Y = cX$ has distribution $\mathcal{N}(c\mu, c^2\sigma^2)$

- If $Z \sim \mathcal{N}(0,1)$, then $\mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$. If $X \sim \mathcal{N}(\mu, \sigma^2)$, we can also write $X \stackrel{d}{=} \mu + \sigma Z$, where $\stackrel{d}{=}$ denotes equality in distribution.

- If $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are **independent** then their sum is normal with matching mean and variance:

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

2.5.2 Gamma Distribution

- Let $X \sim \mathsf{Gamma}(\alpha, \beta)$. For every constant c, the random variable $Y = cX$ has distribution $\mathsf{Gamma}(\alpha, \frac{\beta}{c})$ Therefore,

$$X \overset{d}{=} \frac{1}{\beta} Z \quad \text{where } Z \sim \mathsf{Gamma}(\alpha, 1)$$

- If $X \sim \mathsf{Gamma}(\alpha_1, \beta)$ and $Y \sim \mathsf{Gamma}(\alpha_2, \beta)$ are **independent** then their sum is also gamma:

$$X + Y \sim \mathsf{Gamma}(\alpha_1 + \alpha_2, \beta)$$

2.5.3 Binomial Distribution

- The **Binomial Distribution** with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$ is denoted by $\mathsf{Bin}(n, p)$ and has pmf

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, 2, \ldots, n\}$$

- If $X \sim \mathsf{Bin}(n_1, p)$ and $Y \sim \mathsf{Bin}(n_2, p)$ are **independent** then their sum is also binomial

$$X + Y \sim \mathsf{Bin}(n_1 + n_2, p)$$

- If $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} \mathsf{Ber}(p)$, then $Z_1 + \cdots + Z_n \sim \mathsf{Bin}(n, p)$.

2.5.4 Poisson Distribution

- If $X \sim \mathsf{Poisson}(\lambda_1)$ and $Y \sim \mathsf{Poisson}(\lambda_2)$ are **independent** then their sum is also Poisson

$$X + Y \sim \mathsf{Poisson}(\lambda_1 + \lambda_2)$$