III. Continuous Random Variables

Key concepts

Cumulative distribution function, probability density function, expectation, variance, quantiles.

References

Sections 7.1, 7.2, and 7.3 in the textbook.

Cumulative Distribution Function

Definition

The **cumulative distribution function** of continuous random variable X is

$$P_X(x) = \mathrm{Prob}[X \leq x]\,.$$

Its **tail** is

$$\overline{P}_X(x) = \mathrm{Prob}[X > x] = 1 - P_X(x)\,.$$

- From the probability axioms, $P_X(x)$ is monotonically increasing with x and

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

Probability Density Function

Definition

The **probability density function** of continuous random variable X is defined as

$$p_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} P_X(x)\,.$$

- From the fundamental theorem of calculus,

$$P_X(x) = \int_{-\infty}^{x} p_X(\xi)\,\mathrm{d}\xi\,.$$

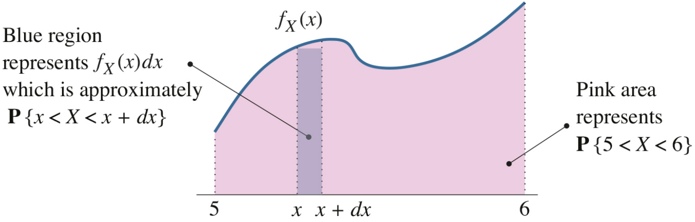
- From the probability axioms, $p_X(x) \geq 0$ for all $x \in \mathbb{R}$ and

$$\int_{-\infty}^{\infty} p_X(x)\,\mathrm{d}x = 1\,.$$

Probability Density Function

- For any $a, b \in \mathbb{R}$ with $a < b$

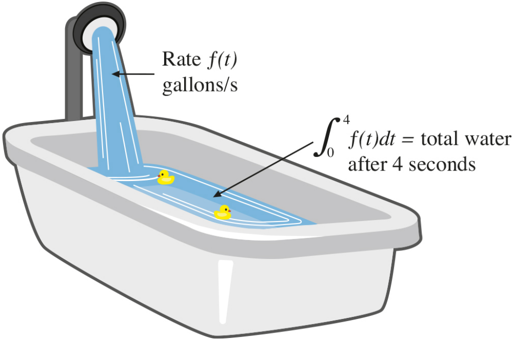
$$\mathrm{Prob}[a < X \leq b] = P_X(b) - P_X(a) = \int_a^b p_X(x)\, \mathrm{d}x\,.$$



- How does $\mathrm{Prob}[a < X < b]$ compare with $\mathrm{Prob}[a \leq X \leq b]$?
- Can $p_X(x)$ be greater than 1?

- Density functions are not unique to probability.

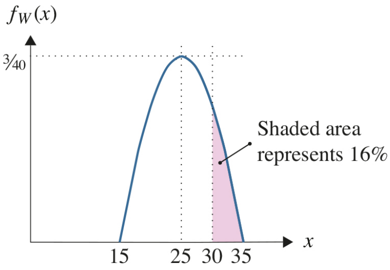
- Assuming that starting at time $t = 0$, a bathtub is filled at a rate of $f(t) = t^2$ gallons per second. How much water is there in the tub after four seconds?



- Assume that the weight (measured in pounds) of two-year-olds is governed by probability density function

$$p_W(x) = \begin{cases} \frac{3}{40} - \frac{3}{4000}(x-25)^2 & \text{if } 15 \leq x \leq 35 \\ 0 & \text{otherwise}\,. \end{cases}$$

- What is the fraction of two-year-olds with a weight exceeding 30 pounds?



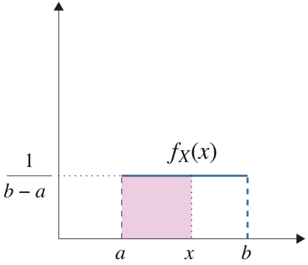
Continuous Random Variables

- Common continuous random variables include:
  - the uniform random variable
  - the exponential random variable
  - the normal random variable
  - the Pareto random variable

Uniform Distribution

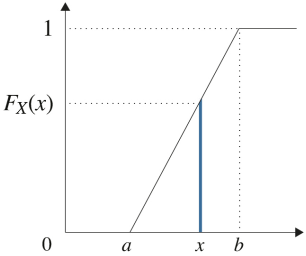
- The **uniform probability distribution** $Uniform(a, b)$, where $a < b$ has probability density

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



- Its cumulative distribution function is

$$\begin{aligned} P_X(x) &= \int_{-\infty}^{x} p_X(\xi)\, \mathrm{d}\xi \\ &= \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } b \leq x\,. \end{cases} \end{aligned}$$



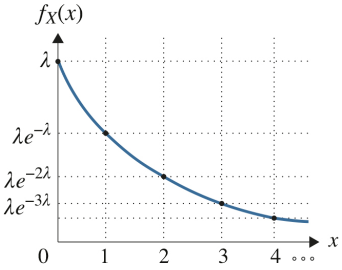
Exponential Distribution

- The **exponential distribution** $\mathit{Exponential}(\lambda)$, where $\lambda > 0$, has probability density function

$$p_X(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise}. \end{cases}$$

- Its cumulative distribution function is

$$\begin{aligned} P_X(x) &= \int_{-\infty}^{x} p_X(\xi)\,\mathrm{d}\xi \\ &= \begin{cases} 1 - \mathrm{e}^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise}. \end{cases} \end{aligned}$$



Expectation

Definition

The **expectation** of continuous random variable X is

$$\mathrm{E}\left[X\right] = \int_{-\infty}^{\infty} x p_X(x)\,\mathrm{d}x\,.$$

- The expectation is the **mean** of the distribution from which X is drawn.
- The expectation of a function f of a continuous random variable X is

$$\mathrm{E}[f(X)] = \int_{-\infty}^{\infty} f(x) p_X(x)\,\mathrm{d}x\,.$$

Expectation

- The expectation of random variable X drawn from uniform distribution $Uniform(a, b)$ is

$$\begin{aligned} \mathrm{E}[X] &= \int_{-\infty}^{\infty} x p_X(x) \, \mathrm{d}x \\ &= \int_a^b \frac{x}{b-a} \, \mathrm{d}x \\ &= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\ &= \frac{a+b}{2} \, . \end{aligned}$$

Expectation

- The expectation of random variable X drawn from exponential distribution $Exponential(\lambda)$ is

$$\begin{aligned}
\mathrm{E}[X] &= \int_{-\infty}^{\infty} x p_X(x)\,\mathrm{d}x \\
&= \lambda \int_0^{\infty} x\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x \\
&= \lambda \left[-\frac{1}{\lambda^2}(1+\lambda x)\,\mathrm{e}^{-\lambda x} \right]_0^{\infty} \\
&= \frac{1}{\lambda}\,.
\end{aligned}$$

Example: The distance from Halifax to Chester is $80\mathrm{km}$. Cyclists cover the distance at constant speeds uniformly distributed between $10\mathrm{km/h}$ and $30\mathrm{km/h}$. What is the expected time needed by a cyclist to travel to Chester?

- Let $X \sim \mathit{Uniform}(10, 30)$ be the speed of the cyclist.
- Then the time required is $T = 80/X$.
- The expectation is

$$\begin{aligned}
\mathrm{E}[T] &= \int_{-\infty}^{\infty} \frac{80}{x} p_X(x)\,\mathrm{d}x \\
&= \frac{1}{20} \int_{10}^{30} \frac{80}{x}\,\mathrm{d}x \\
&= 4\,[\log(x)]_{10}^{30} \\
&= 4(\log(30) - \log(10)) \\
&= 4\log(3)\,.
\end{aligned}$$

Variance

Definition

The **variance** of continuous random variable X is

$$\mathrm{Var}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] .$$

- Analogous to the discrete case,

$$\begin{aligned}
\mathrm{Var}[X] &= \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 \, p_X(x) \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \left(x^2 - 2x\mathrm{E}[X] + \mathrm{E}[X]^2\right) p_X(x) \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} x^2 p_X(x) \, \mathrm{d}x - 2\mathrm{E}[X] \int_{-\infty}^{\infty} x p_X(x) \, \mathrm{d}x + \mathrm{E}[X]^2 \int_{-\infty}^{\infty} p_X(x) \, \mathrm{d}x \\
&= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 .
\end{aligned}$$

Variance

- The variance of random variable X drawn from uniform distribution $Uniform(a, b)$ is

$$\begin{aligned}
\mathrm{Var}[X] &= \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 \, p_X(x) \, \mathrm{d}x \\
&= \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2} \right)^2 \mathrm{d}x \\
&= \frac{1}{b-a} \left[\frac{1}{3} \left(x - \frac{a+b}{2} \right)^3 \right]_a^b \\
&= \frac{(b-a)^2}{12} .
\end{aligned}$$

Variance

- The variance of random variable X drawn from exponential distribution $Exponential(\lambda)$ is

$$\begin{aligned}
\mathrm{Var}[X] &= \int_{-\infty}^{\infty} x^2 p_X(x)\,\mathrm{d}x - \mathrm{E}[X]^2 \\
&= \lambda \int_0^{\infty} x^2\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x - \frac{1}{\lambda^2} \\
&= \lambda \left[-\frac{1}{\lambda^3}\left(\lambda^2 x^2 + 2\lambda x + 2\right)\mathrm{e}^{-\lambda x}\right]_0^{\infty} - \frac{1}{\lambda^2} \\
&= \frac{1}{\lambda^2}\,.
\end{aligned}$$

Quantiles

- Let X be a continuous random variable. The **quantile function** $Q_X : [0,1] \to \mathbb{R}$ is defined as
$$Q_X(p) = P_X^{-1}(x)\,.$$
- It associates with probability p the value $x \in \mathbb{R}$ for which $\mathrm{Prob}[X \leq x] = p$.
- The **quartiles** are the values of the quantile function at $p \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$.
- The second quartile is the **median**.

Example: Plot the quantile functions of the uniform and exponential random variables. Then determine their medians.

Summary

- Probability density functions of continuous random variables are the first-order derivatives of their cumulative distribution functions.
- Probabilities can be computed from probability density functions by means of integration.
- Expectation and variance are defined analogously to the discrete case. They can be obtained from probability density functions through integration.
- The quantile function is the inverse of the cumulative distribution function.

Conditioning on a Random Variable

Key concepts

Conditional probability density, conditional expectation.

References

Sections 7.4 and 7.5 in the textbook.

Conditioning on a Random Variable

- Let E be an event and X be a continuous random variable. Then, from the Law of Total Probability,

$$\mathrm{Prob}[E] = \int_{-\infty}^{\infty} p_X(x)\,\mathrm{Prob}[E \mid X = x]\,\mathrm{d}x\,.$$

- Notice that we are conditioning on a zero-probability event.

DALHOUSIE UNIVERSITY

Example: Let X be exponentially distributed with rate $\lambda = 1$ and let E be the event that $X > a$, where $a > 0$. Then

$$
\begin{aligned}
\mathrm{Prob}[E] &= \int_{-\infty}^{\infty} p_X(x)\,\mathrm{Prob}[X > a \mid X = x]\,\mathrm{d}x \\
&= \int_{a}^{\infty} p_X(x)\,\mathrm{d}x \\
&= \int_{a}^{\infty} \mathrm{e}^{-x}\,\mathrm{d}x \\
&= \left[-\mathrm{e}^{-x}\right]_a^{\infty} \\
&= \mathrm{e}^{-a}\,.
\end{aligned}
$$

Conditioning on a Random Variable

Example: Assume that we have a coin with probability X of heads, where $X \sim \mathit{Uniform}(0,1)$. Let E be the event that the next n flips are all heads. Then

$$\begin{aligned}
\mathrm{Prob}[E] &= \int_{-\infty}^{\infty} p_X(x)\,\mathrm{Prob}[E \mid X = x]\,\mathrm{d}x \\
&= \int_0^1 x^n\,\mathrm{d}x \\
&= \left[\frac{x^{n+1}}{n+1}\right]_0^1 \\
&= \frac{1}{n+1}.
\end{aligned}$$

- In MATLAB:

```
M = 1000000;
n = 7;
x = rand(1, M);
sum(all(rand(n, M)<x))/M
```

Conditional Probability Density

- Let E be an event with $\mathrm{Prob}[E] > 0$ and let

$$P_{X|E}(x) = \mathrm{Prob}[X \leq x \mid E] = \frac{\mathrm{Prob}[X \leq x \cap E]}{\mathrm{Prob}[E]}\,.$$

Definition

The **conditional probability density of** X **given event** E is

$$p_{X|E}(x) = \frac{\mathrm{d}}{\mathrm{d}x} P_{X|E}(x)\,.$$

- Note that

$$\int_{-\infty}^{\infty} p_{X|E}(x)\,\mathrm{d}x = 1\,.$$

Conditional Probability Density

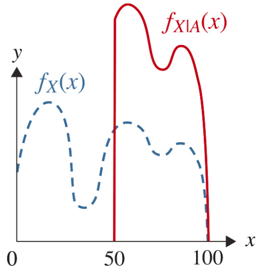
- Bayes' Law then reads

$$\mathrm{Prob}[E]\, p_{X|E}(x) = p_X(x)\, \mathrm{Prob}[E \,|\, X = x]$$

and thus

$$p_{X|E}(x) = \frac{p_X(x)\, \mathrm{Prob}[E \,|\, X = x]}{\mathrm{Prob}[E]}\,.$$

Example: Conditioning on the event that $X \geq 50$.



Example: Let X be exponentially distributed with rate $\lambda = 1$ and let E be the event that $X > a$. The conditional probability density of X given E is

$$p_{X|E}(x) = \frac{p_X(x)\,\mathrm{Prob}[E \mid X = x]}{\mathrm{Prob}[E]} = \begin{cases} \mathrm{e}^{a-x} & \text{if } x > a \\ 0 & \text{otherwise}\,. \end{cases}$$

Conditional Probability Density

Example: Assume that we have a coin with probability X of heads, where $X \sim Uniform(0,1)$. Let E be the event that the next n flips are all heads. Then

$$p_{X|E}(x) = \frac{p_X(x)\,\mathrm{Prob}[E \mid X = x]}{\mathrm{Prob}[E]} = \begin{cases} (n+1)x^n & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise}\,. \end{cases}$$

Conditional Expectation

Definition

Let X be a continuous random variable with probability density function $p_X(x)$. Let E be an event with $\mathrm{Prob}[E] > 0$. Then the **conditional expectation of X given event E** is

$$\mathrm{E}[X|E] = \int_{-\infty}^{\infty} x p_{X|E}(x)\,\mathrm{d}x\,.$$

Example: Let X be exponentially distributed with rate $\lambda = 1$ and let E be the event that $X > a$. Then the conditional expectation of X given E is

$$\begin{aligned}
\mathrm{E}[X|E] &= \int_{-\infty}^{\infty} x p_{X|E}(x)\,\mathrm{d}x \\
&= \int_{a}^{\infty} x\,\mathrm{e}^{a-x}\,\mathrm{d}x \\
&= \left[-(1+x)\,\mathrm{e}^{a-x}\right]_a^{\infty} \\
&= 1 + a\,.
\end{aligned}$$

DALHOUSIE UNIVERSITY

Example: Assume that we have a coin with probability X of heads, where $X \sim Uniform(0,1)$. Let E be the event that the next n flips are all heads. Then the conditional expected bias is

$$\begin{aligned}
\mathrm{E}[X|E] &= \int_{-\infty}^{\infty} x p_{X|E}(x)\,\mathrm{d}x \\
&= (n+1)\int_0^1 x^{n+1}\,\mathrm{d}x \\
&= (n+1)\left[\frac{x^{n+2}}{n+2}\right]_0^1 \\
&= \frac{n+1}{n+2}\,.
\end{aligned}$$

Notice that the answer depends crucially on the initial assumption that $X \sim Uniform(0,1)$, which is referred to as a **prior**.

Example: Runtime Bins

Example: At a supercomputing centre, jobs are grouped into different bins based on their size. Job sizes are exponentially distributed with an expectation of 1000 CPU hours. All jobs that require less than 500 CPU hours (and only those) are sent to bin 1.

- What is the probability that a job is sent to bin 1?
- For jobs in bin 1, what is the probability that they require less than 200 CPU hours?
- What is the conditional probability density of all jobs in bin 1?
- What is the expected CPU time requirement for jobs in bin 1?

Conditioning on a Random Variable

Summary

- Probabilities may be computed by conditioning on a random variable.
- Conditional probability densities are probability density functions conditioned on an event.
- Conditional expectations can be computed from conditional probability densities.

Key concepts

Joint probability density, marginal distribution, independence.

References

Chapter 8 in the textbook.

Joint Distribution

DALHOUSIE UNIVERSITY

Definition

The **joint cumulative distribution function** of continuous random variables X and Y is

$$P_{X,Y}(x,y) = \mathrm{Prob}[X \leq x \cap Y \leq y]\,.$$

Their **joint probability density function** is

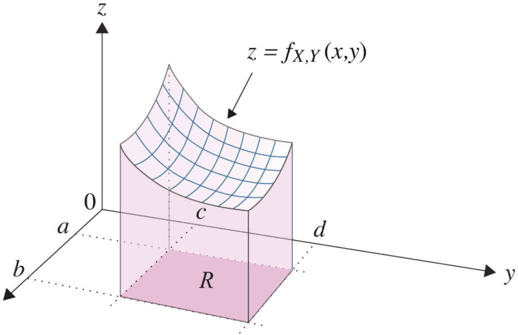
$$p_{X,Y}(x,y) = \frac{\partial^2 P_{X,Y}(x,y)}{\partial x \partial y}\,.$$

- From the probability axioms, $p_{X,Y}(x,y) \geq 0$ for all $x, y \in \mathbb{R}$ and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x = 1\,.$$

- Probabilities are obtained through integration: for any $a_1, a_2, b_1, b_2 \in \mathbb{R}$ with $a_1 < b_1$ and $a_2 < b_2$

$$\mathrm{Prob}[a_1 \leq X \leq b_1 \cap a_2 \leq Y \leq b_2] = \int_{a_1}^{b_1} \int_{a_2}^{b_2} p_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x\,.$$



Example: If $p_{X,Y}(x,y)$ is the joint probability density of weight X (measured in pounds) and height Y (measured in centimetres) of two-year-olds, then the fraction of two-year-olds with a weight exceeding 30 pounds and a height less than 80cm is

$$\int_{30}^{\infty} \int_{0}^{80} p_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x\,.$$

Marginal Probability Density

Definition

The **marginal probability densities** of continuous random variables X and Y with joint density $p_{X,Y}(x,y)$ are

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,\mathrm{d}y$$

and

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,\mathrm{d}x\,.$$

- The marginal densities above are univariate probability densities with

$$\mathrm{Prob}[X \leq x] = \int_{-\infty}^{x} p_X(\xi)\,\mathrm{d}\xi \quad \text{and} \quad \mathrm{Prob}[Y \leq y] = \int_{-\infty}^{y} p_Y(\upsilon)\,\mathrm{d}\upsilon\,.$$

Definition

Continuous random variables X and Y with joint probability density $p_{X,Y}(x, y)$ are **independent** if and only if for all $x, y \in \mathbb{R}$,

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \, .$$

- Are X and Y independent if their joint density is

$$p_{X,Y}(x, y) = \begin{cases} 4xy & \text{if } 0 \leq x, y \leq 1 \\ 0 & \text{otherwise}\,? \end{cases}$$

- How about X and Y with joint density

$$p_{X,Y}(x, y) = \begin{cases} x + y & \text{if } 0 \leq x, y \leq 1 \\ 0 & \text{otherwise}\,? \end{cases}$$

Example: Assume that you have two servers.

- The time until server 1 crashes is $X \sim Exponential(\lambda)$.
- The time until server 2 crashes is $Y \sim Exponential(\mu)$.

Assuming independence, what is the probability that server 1 crashes before server 2?

- Integration yields

$$
\begin{aligned}
\mathrm{Prob}[X < Y] &= \int_0^\infty \int_x^\infty p_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_0^\infty p_X(x) \int_x^\infty p_Y(y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \lambda\mu \int_0^\infty \mathrm{e}^{-\lambda x} \int_x^\infty \mathrm{e}^{-\mu y}\,\mathrm{d}y\,\mathrm{d}x \\
&= \lambda\mu \int_0^\infty \mathrm{e}^{-\lambda x} \left[-\frac{1}{\mu}\mathrm{e}^{-\mu y}\right]_x^\infty \mathrm{d}x \\
&= \lambda \int_0^\infty \mathrm{e}^{-(\mu+\lambda)x}\,\mathrm{d}x \\
&= \lambda \left[-\frac{1}{\mu+\lambda}\mathrm{e}^{-(\mu+\lambda)x}\right]_0^\infty \\
&= \frac{\lambda}{\mu+\lambda}\,.
\end{aligned}
$$

Conditional Density Functions

Definition

Let X and Y be continuous random variables with joint probability density function $p_{X,Y}(x, y)$. Then the **conditional probability density function of X given Y** is

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \,.$$

- For all $y \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} p_{X|Y}(x|y) \, \mathrm{d}x = 1 \,.$$

Conditional Density Functions

- Bayes' Law: for all $x, y \in \mathbb{R}$,

$$p_Y(y)p_{X|Y}(x|y) = p_X(x)p_{Y|X}(y|x)$$

- Law of Total Probability:

$$p_X(x) = \int_{-\infty}^{\infty} p_Y(y)p_{X|Y}(x|y)\,\mathrm{d}y$$

Example: Assume that you have two servers.

- The time until server 1 crashes is $X \sim Exponential(\lambda)$.
- The time until server 2 crashes is $Y \sim Exponential(\mu)$.

Assuming independence, what it the probability that server 1 crashes before server 2?

- By conditioning on the time that server 1 crashes:

$$\begin{aligned}
\mathrm{Prob}[X < Y] &= \int_0^\infty p_X(x)\ \mathrm{Prob}[X < Y \mid X = x]\,\mathrm{d}x \\
&= \int_0^\infty p_X(x)\ \mathrm{Prob}[x < Y]\,\mathrm{d}x \\
&= \lambda \int_0^\infty \mathrm{e}^{-\mu x}\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x \\
&= \lambda \left[-\frac{\mathrm{e}^{-(\mu+\lambda)x}}{\mu+\lambda}\right]_0^\infty \\
&= \frac{\lambda}{\mu+\lambda}\,.
\end{aligned}$$

Example: Homework

- Assume that the joint probability density between the time X (measured in days) that an assignment is handed in early and the grade Y (as a percentage) is

$$p_{X,Y}(x,y) = \begin{cases} 0.9xy^2 + 0.2 & \text{if } 0 \leq x \leq 2 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise}. \end{cases}$$

  - What is the probability that a random student earns a grade of at least 0.5?
  - What is the probability for student who submits less than a day before the deadline to earn a grade of at least 0.5?

Lemma

Let X and Y be independent random variables with probability densities $p_X(x)$ and $p_Y(y)$, respectively. Then the probability density of $Z = X + Y$ is

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(x) p_Y(z - x) \, \mathrm{d}x \, .$$

Proof: The cumulative distribution function of Z is

$$\begin{aligned} P_Z(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} p_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} p_X(x) \int_{-\infty}^{z-x} p_Y(y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} p_X(x) P_Y(z-x)\,\mathrm{d}x\,. \end{aligned}$$

Differentiation yields

$$\begin{aligned} p_Z(z) &= \frac{\mathrm{d}}{\mathrm{d}z} P_Z(z) \\ &= \frac{\mathrm{d}}{\mathrm{d}z} \int_{-\infty}^{\infty} p_X(x) P_Y(z-x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} p_X(x) \left(\frac{\mathrm{d}}{\mathrm{d}z} P_Y(z-x) \right) \mathrm{d}x \\ &= \int_{-\infty}^{\infty} p_X(x) p_Y(z-x)\,\mathrm{d}x\,. \qquad \square \end{aligned}$$

- If X and Y are continuous random variables with joint probability density function $p_{X,Y}(x, y)$, then for any function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathrm{E}[f(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) p_{X,Y}(x, y) \, \mathrm{d}y \, \mathrm{d}x \, .$$

DALHOUSIE UNIVERSITY

Lemma

If X and Y are continuous random variables and $a, b \in \mathbb{R}$, then

$$\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]\,.$$

Proof:

$$\begin{aligned}
\mathrm{E}[aX + bY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax + by)p_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= a\int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x + b\int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= a\int_{-\infty}^{\infty} x p_X(x)\,\mathrm{d}x + b\int_{-\infty}^{\infty} y p_Y(y)\,\mathrm{d}y \\
&= a\mathrm{E}[X] + b\mathrm{E}[Y]\,. \qquad \square
\end{aligned}$$

Expectation

Lemma

If X and Y are independent, continuous random variables, then

$$\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]\,.$$

Proof:

$$\begin{aligned}
\mathrm{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy p_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy p_X(x) p_Y(y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \left(\int_{-\infty}^{\infty} x p_X(x)\,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} y p_Y(y)\,\mathrm{d}y\right) \\
&= \mathrm{E}[X]\,\mathrm{E}[Y]\,. \qquad \square
\end{aligned}$$

Lemma

If X and Y are independent, continuous random variables and $a, b \in \mathbb{R}$, then

$$\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y]\,.$$

Proof:

$$\begin{aligned}
\mathrm{Var}[aX + bY] &= \mathrm{E}\big[(aX + bY)^2\big] - \mathrm{E}[aX + bY]^2 \\
&= a^2\mathrm{E}\big[X^2\big] + 2ab\mathrm{E}[XY] + b^2\mathrm{E}\big[Y^2\big] \\
&\qquad - a^2\mathrm{E}[X]^2 - 2ab\mathrm{E}[X]\mathrm{E}[Y] - b^2\mathrm{E}[Y]^2 \\
&= a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y]\,. \qquad \square
\end{aligned}$$

Conditional Expectation

- The expectation of X conditional on $Y = y$ is

$$\mathrm{E}[X|Y = y] = \int_{-\infty}^{\infty} x p_{X|Y}(x|y)\,\mathrm{d}x\,.$$

- The expectation of random variable X can be derived by conditioning on random variable Y:

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} \mathrm{E}[X \,|\, Y = y]\, p_Y(y)\,\mathrm{d}y\,.$$

Example: Homework

- Assume that the joint probability density between the time X (measured in days) that an assignment is handed in early and the grade Y (as a percentage) is

$$p_{X,Y}(x,y) = \begin{cases} 0.9xy^2 + 0.2 & \text{if } 0 \leq x < 2 \text{ and } 0 \leq y < 1 \\ 0 & \text{otherwise} \,. \end{cases}$$

  - What is the expected grade of a student submitting their homework exactly when it is due?
  - What is the expected grade of a student submitting their homework at least a day before the due date?

Summary

- Multiple continuous random variables are governed by joint probability density functions.
- Joint density functions of independent random variables are products of their marginal density functions.
- The expectation of a linear combination of random variables is a linear combination of their expectations.
- The variance of a linear combination of random variables is a linear combination of their variances only if the random variables are independent.
- Conditioning can simplify the computation of probabilities and expectations.

Key concepts

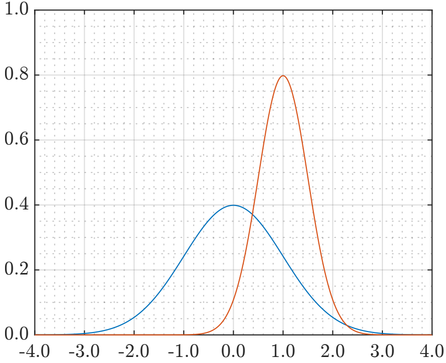
Normal distribution, Central Limit Theorem.

References

Chapter 9 in the textbook.

- The normal probability distribution $Normal(\mu, \sigma^2)$ has probability density function

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



Probability density functions of normally distributed random variables.

- The expected value of the normal distribution is

$$
\begin{aligned}
\mathrm{E}[X] &= \int_{-\infty}^{\infty} x p_X(x)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} (\mu+\sigma u)\,\mathrm{e}^{-u^2/2}\,\mathrm{d}u \\
&= \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-u^2/2}\,\mathrm{d}u + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u\,\mathrm{e}^{-u^2/2}\,\mathrm{d}u \\
&= \mu\,,
\end{aligned}
$$

where substitution $u = (x-\mu)/\sigma$ and the facts that $\int_{-\infty}^{\infty} \mathrm{e}^{-u^2/2}\,\mathrm{d}u = \sqrt{2\pi}$ and $\int_{-\infty}^{\infty} u\,\mathrm{e}^{-u^2/2}\,\mathrm{d}u = 0$ have been used.

- The variance of the normal distribution is

$$
\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 p_X(x)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 \mathrm{e}^{-u^2/2}\,\mathrm{d}u \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left[-u\,\mathrm{e}^{-u^2/2}\right]_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-u^2/2}\,\mathrm{d}u \\
&= \sigma^2
\end{aligned}
$$

where substitution $u = (x-\mu)/\sigma$ and integration by parts have been used.

- The cumulative distribution function

$$P_X(x) = \int_{-\infty}^{x} p_X(\xi)\,\mathrm{d}\xi$$

of normally distributed random variable X cannot be expressed in closed form.

- The normal distribution with parameters $\mu = 0$ and $\sigma = 1$ is the **standard normal distribution**.
- We write

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\xi^2/2}\,\mathrm{d}\xi$$

for the cumulative distribution function of a standard normal random variable.

Lemma

If $X \sim Normal(0,1)$, then for $\sigma > 0$, $Y = \mu + \sigma X \sim Normal(\mu, \sigma^2)$.

Proof: From

$$P_Y(y) = \text{Prob}[Y \le y] = \text{Prob}[\mu + \sigma X \le y] = \text{Prob}[X \le (y-\mu)/\sigma] = \Phi((y-\mu)/\sigma)$$

it follows that the probability density of Y is

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}y} P_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} \Phi\left(\frac{y-\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) .\end{aligned}$$

□

Normal Distribution

- MATLAB command `randn` samples from the standard normal distribution.
- To generate a sample from $Normal(\mu, \sigma^2)$, compute $\mu + \sigma X$, where $X \sim Normal(0, 1)$.
- MATLAB command `normcdf` numerically approximates the cumulative distribution function of the standard normal distribution.
- To obtain $P_X(x)$ where $X \sim Normal(\mu, \sigma^2)$, compute $\Phi((x - \mu)/\sigma)$.

Normal Distribution

Lemma

If

$$X \sim Normal(\mu_X, \sigma_X^2)$$

and independently

$$Y \sim Normal(\mu_Y, \sigma_Y^2),$$

then

$$Z = X + Y \sim Normal(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

Proof: We could show this with the result on the sum of independent random variables shown above, but it will be easier after having introduced Laplace transforms below.

- By induction, the sum of any number of normal random variables is normally distributed.

Central Limit Theorem

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables, each with mean μ and finite variance σ^2, and let

$$S_n = \sum_{k=1}^{n} X_k \, .$$

Then for every $z \in \mathbb{R}$,

$$\lim_{n \to \infty} \mathrm{Prob}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right] = \Phi(z) \, .$$

- Note that $\mathrm{E}[S_n] = n\mu$ and $\mathrm{Var}[S_n] = n\sigma^2$.
- Random variables X_k may be either discrete or continuous.

Central Limit Theorem

- The Central Limit Theorem can be used to approximate the cumulative effect of many additive terms.
- It can be proven using Laplace transforms.

Example: Consider n independent random variables $X_k \sim Uniform(a, b)$, $k = 1, 2, \ldots, n$.

- Then $\mathrm{E}[X_k] = (a + b)/2 = \mu$ and $\mathrm{Var}[X_k] = (b - a)^2/12 = \sigma^2$.
- The probability that $c < S_n < d$ is

$$\begin{aligned}\mathrm{Prob}[c < S_n < d] &= \mathrm{Prob}\left[\frac{c - n\mu}{\sqrt{n}\sigma} < \frac{S_n - n\mu}{\sqrt{n}\sigma} < \frac{d - n\mu}{\sqrt{n}\sigma}\right] \\ &\approx \Phi\left(\frac{d - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{c - n\mu}{\sqrt{n}\sigma}\right).\end{aligned}$$

Central Limit Theorem

```
function [P, Q] = CLTapprox(n, a, b, c, d)
  mu = (a+b)/2;
  sigma = (b-a)/sqrt(12);
  P = normcdf((d-n*mu)/sqrt(n)/sigma) ...
        - normcdf((c-n*mu)/sqrt(n)/sigma);

  M = 1000000;
  S = sum(a+(b-a)*rand(n, M));
  Q = sum(c<S & S<d)/M;

  h = histogram(S);
  hold on

  x = linspace(h.BinLimits(1), h.BinLimits(2), h.NumBins);
  plot(x, h.BinWidth*M ...
            * exp(-(x-n*mu).^2/sigma^2/n/2)/sqrt(2*pi)/sigma/sqrt(n))
  hold off
end
```

Summary

- Any linear function applied a normal random variable results in a normal random variable.
- The sum of independent normal random variables is normally distributed.
- The Central Limit Theorem assures us that the average of multiple i.i.d. random variables becomes normal in the limit of infinitely many variables.

Pareto Distribution

Key concepts

Memorylessness, failure rate, Pareto distribution.

References

Section 7.2 and Chapter 10 in the textbook.

Memorylessness

Definition

Random variable X is **memoryless** if and only if for all $x, t \geq 0$

$$\mathrm{Prob}[X > x] = \mathrm{Prob}[X > t + x \,|\, X > t]\,.$$

Memorylessness

- Time until winning the lottery is memoryless.

  Proof: Time X (measured in the number of tickets bought) until winning the lottery is geometrically distributed:

$$\text{Prob}[X > k] = \overline{P}_X(k) = (1-p)^k ,$$

  where p is the probability of picking the right numbers and $k \geq 0$. Thus, for any $t \geq 0$,

$$\begin{aligned}\text{Prob}[X > t+k \,|\, X > t] &= \frac{\text{Prob}[X > t+k \cap X > t]}{\text{Prob}[X > t]} \\ &= \frac{\text{Prob}[X > t+k]}{\text{Prob}[X > t]} \\ &= \frac{(1-p)^{t+k}}{(1-p)^t} \\ &= (1-p)^k \\ &= \text{Prob}[X > k] \,. \end{aligned} \qquad \square$$

- The exponential distribution is memoryless.

Proof: Let X be exponentially distributed. From the cumulative distribution function of the exponential distribution, for $x \geq 0$

$$\text{Prob}[X > x] = \overline{P}_X(x) = \mathrm{e}^{-\lambda x}\,.$$

Thus, for any $t \geq 0$,

$$\begin{aligned}
\text{Prob}[X > t + x \,|\, X > t] &= \frac{\text{Prob}[X > t + x \cap X > t]}{\text{Prob}[X > t]} \\
&= \frac{\text{Prob}[X > t + x]}{\text{Prob}[X > t]} \\
&= \frac{\mathrm{e}^{-\lambda(t+x)}}{\mathrm{e}^{-\lambda t}} \\
&= \mathrm{e}^{-\lambda x} \\
&= \text{Prob}[X > x]\,. \qquad \square
\end{aligned}$$

Memorylessness

- Assume that the lifetime X (measured in years) of the naked mole-rat is exponentially distributed with rate $\lambda = 1$.
- If a naked mole-rat is four years old, then the probability of it surviving at least one more year is

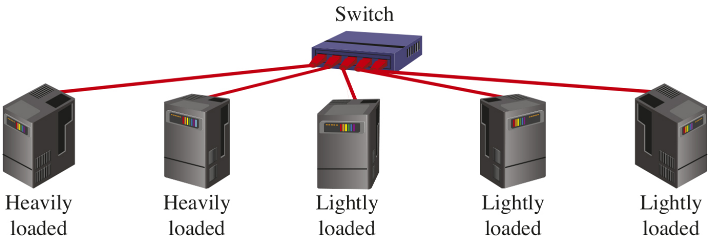
$$\mathrm{Prob}[X > 5 \mid X > 4] = \frac{\mathrm{Prob}[X > 5]}{\mathrm{Prob}[X > 4]} = \frac{\mathrm{e}^{-5}}{\mathrm{e}^{-4}} = \mathrm{e}^{-1}\,.$$

- If a naked mole-rat is 27 years old, then the probability of it surviving at least one more year is

$$\mathrm{Prob}[X > 28 \mid X > 27] = \frac{\mathrm{Prob}[X > 28]}{\mathrm{Prob}[X > 27]} = \frac{\mathrm{e}^{-28}}{\mathrm{e}^{-27}} = \mathrm{e}^{-1}\,.$$

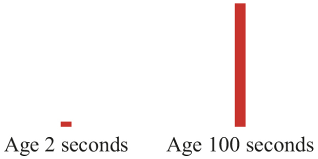
- A post office has two clerks. All service times are exponentially distributed with rate λ.
- When customer C walks in, customer A is served by one clerk and customer B by the other.
- What is the probability that customer C is the last to leave?

- Consider the task of CPU load balancing.
- Jobs can be migrated after they have started running (**preemptive migration**) or only before.
- Migrating preemptively is relatively costly.



Pareto Distribution

- A job's **size** is its total CPU requirement.
- A job's **age** is its total CPU usage so far.
- A job's **remaining size** is the difference between the two.



- Only a job's age is known.
- A job's remaining size determines whether preemptive migration is worth it.
- Let X be a job's size. It would be useful to know

$$\mathrm{Prob}[X > x + a \mid X > a]\,.$$

Pareto Distribution

- For a distribution that is memoryless, the remaining size of a job is independent of its age.
- A job's expected remaining size may increase with age ("decreasing failure rate"), or it may decrease ("increasing failure rate").
- Can you think of real-world examples with increasing/decreasing failure rate?

Definition

The **failure rate** of continuous random variable X is

$$r_X(t) = \frac{p_X(t)}{\overline{P}_X(t)}\,.$$

- Consider the probability that the failure of an item with age t is imminent:

$$\begin{aligned}\mathrm{Prob}[X \in [t, t + \mathrm{d}t) \mid X \geq t] &= \frac{\mathrm{Prob}[X \in [t, t + \mathrm{d}t)]}{\mathrm{Prob}[X \geq t]} \\ &= \frac{p_X(t)\,\mathrm{d}t}{\overline{P}_X(t)} \\ &= r_X(t)\,\mathrm{d}t\end{aligned}$$

$\Rightarrow$ $r_X(t)$ is the instantaneous failure rate of an item of age t.

- The failure rate of $X \sim \mathit{Exponential}(\lambda)$ is

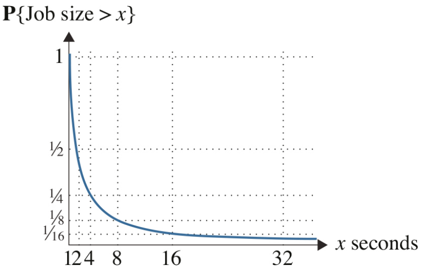
$$
\begin{aligned}
r_X(t) &= \frac{\lambda\,\mathrm{e}^{-\lambda t}}{\mathrm{e}^{-\lambda t}} \\
&= \lambda\,.
\end{aligned}
$$

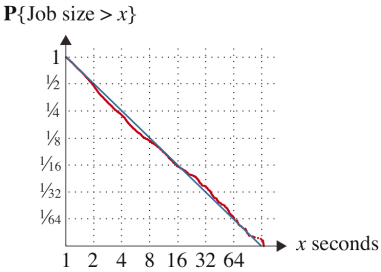
- It can be shown that the exponential distribution is the only continuous distribution with a constant failure rate.

Pareto Distribution

- Preemptive migration is worth it only if failure rates of jobs decrease with increasing age.
- Empirically observed tails of the size X of UNIX jobs for $x \geq 1$ are well approximated by

$$\overline{P}_X(x) = \frac{1}{x}.$$





Pareto Distribution

- The **Pareto distribution** $Pareto(\alpha)$ $(0 < \alpha < 2)$ has tail

$$\overline{P}_X(x) = \begin{cases} x^{-\alpha} & \text{if } x \geq 1 \\ 1 & \text{otherwise} \,. \end{cases}$$

- Its cumulative distribution function is

$$P_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 - x^{-\alpha} & \text{if } x \geq 1 \,. \end{cases}$$

- Its probability density is

$$p_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \alpha x^{-\alpha-1} & \text{if } x \geq 1 \,. \end{cases}$$

Pareto Distribution

- What is the mean of the Pareto distribution?
- The failure rate of the Pareto distribution for $x \geq 1$ is

$$r_X(x) = \frac{\alpha}{x}$$

  and thus decreasing with x.
- The Pareto distribution has a **heavy tail**. Preemptive migration may be worth it.
- Other quantities that are (within bounds) well modelled by Pareto distributions include file sizes, internet node degrees, IP flow durations, ...

Pareto Distribution

Distribution	p.d.f. $f_X(x)$	Mean	Variance
Exp(λ)	$f_X(x) = \lambda e^{-\lambda x}, x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Uniform(a, b)	$f_X(x) = \frac{1}{b-a}$, if $a \leq x \leq b$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Pareto(α), $0 < \alpha < 2$	$f_X(x) = \alpha x^{-\alpha-1}$, if $x > 1$	$\begin{cases} \infty & \text{if } \alpha \leq 1 \\ \frac{\alpha}{\alpha-1} & \text{if } \alpha > 1 \end{cases}$	∞
Normal(μ, σ^2)	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $-\infty < x < \infty$	μ	σ^2

Summary

- Failure rates quantify the probability that an item's failure is imminent.
- Failure rates may increase or decrease with age.
- The exponential distribution is memoryless and has a constant failure rate.
- The Pareto distribution exhibits failure rates that decrease with age.
- Pareto distributions are characterized by heavy tails, may have no finite expectation, and have no finite variance.

Moments and Laplace Transform

Key concepts

Moments, central moments, Laplace transform.

References

Chapter 11 in the textbook.

Moments

Definition

The k**th moment** of continuous random variable X is

$$\mathrm{E}\left[X^k\right] = \int_{-\infty}^{\infty} x^k p_X(x)\,\mathrm{d}x\,.$$

- The expectation is the first moment.

Central Moments

Definition

The k**th central moment** of continuous random variable X is

$$\mathrm{E}\left[(X - \mathrm{E}[X])^k\right] .$$

- The variance is the second central moment.
- Moments can be computed from central moments and vice versa. E.g.,

$$\mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2$$

- The third and fourth central moments capture skewness and kurtosis.

- The Laplace transform allows easily obtaining moments of nonnegative, continuous random variables.

Definition

The **Laplace transform**, $\mathcal{L}_X(z)$, of nonnegative, continuous random variable X is defined for $z \geq 0$ as

$$\mathcal{L}_X(z) = \mathrm{E}\left[\mathrm{e}^{-zX}\right] = \int_0^\infty \mathrm{e}^{-zx}\, p_X(x)\,\mathrm{d}x\,.$$

- The value of the Laplace transform at $z = 0$ is 1.

- The Laplace transform converges for all $z \geq 0$.

Lemma

For any nonnegative, continuous random variable X and $z \geq 0$, $0 \leq \mathcal{L}_X(z) \leq 1$.

Proof:

- As $z \geq 0$, for $x \geq 0$ it follows that $0 < \mathrm{e}^{-zx} \leq 1$.
- The lemma follows immediately from

$$\int_0^\infty p_X(x)\,\mathrm{d}x = 1\,. \qquad \square$$

Laplace Transform

Theorem

Let X be a nonnegative, continuous random variable and $n \in \mathbb{N}$. Then

$$\frac{\mathrm{d}^n}{\mathrm{d}z^n}\mathcal{L}_X(z)\bigg|_{z=0} = (-1)^n \mathrm{E}\left[X^k\right] .$$

- Using the power series expansion of the exponential function it follows that

$$\begin{aligned}
\mathcal{L}_X(z) &= \int_0^\infty \mathrm{e}^{-zx}\, p_X(x)\,\mathrm{d}x \\
&= \sum_{k=0}^\infty \frac{(-z)^k}{k!} \int_0^\infty x^k p_X(x)\,\mathrm{d}x \\
&= \sum_{k=0}^\infty \frac{(-z)^k}{k!} \mathrm{E}\left[X^k\right] .
\end{aligned}$$

- Consider $n = 1$. Then

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathcal{L}_X(z) = \frac{\mathrm{d}}{\mathrm{d}z}\sum_{k=0}^{\infty}\frac{(-z)^k}{k!}\mathrm{E}\left[X^k\right]$$
$$= -\sum_{k=1}^{\infty}\frac{(-z)^{k-1}}{(k-1)!}\mathrm{E}\left[X^k\right] .$$

- Thus,

$$\left.\frac{\mathrm{d}}{\mathrm{d}z}\mathcal{L}_X(z)\right|_{z=0} = -\mathrm{E}[X] .$$

$\Rightarrow$ The first-order derivative is the negative of the first moment.

Laplace Transform

- Consider $n = 2$. Then

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\mathcal{L}_X(z) = -\frac{\mathrm{d}}{\mathrm{d}z}\sum_{k=1}^{\infty}\frac{(-z)^{k-1}}{(k-1)!}\mathrm{E}\left[X^k\right]$$
$$= \sum_{k=2}^{\infty}\frac{(-z)^{k-2}}{(k-2)!}\mathrm{E}\left[X^k\right] .$$

- Thus,

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}z^2}\mathcal{L}_X(z)\right|_{z=0} = \mathrm{E}\left[X^2\right] .$$

$\Rightarrow$ The second-order derivative is the second moment.

Laplace Transform

- Consider $n = 3$. Then

$$\frac{\mathrm{d}^3}{\mathrm{d}z^3}\mathcal{L}_X(z) = \frac{\mathrm{d}}{\mathrm{d}z}\sum_{k=2}^{\infty}\frac{(-z)^{k-2}}{(k-2)!}\mathrm{E}\left[X^k\right]$$

$$= -\sum_{k=3}^{\infty}\frac{(-z)^{k-3}}{(k-3)!}\mathrm{E}\left[X^k\right] .$$

- Thus,

$$\left.\frac{\mathrm{d}^3}{\mathrm{d}z^3}\mathcal{L}_X(z)\right|_{z=0} = -\mathrm{E}\left[X^3\right] .$$

$\Rightarrow$ The third-order derivative is the negative of the third moment.

- The Laplace transform of $X \sim \mathit{Exponential}(\lambda)$ is

$$\begin{aligned}\mathcal{L}_X(z) &= \lambda \int_0^\infty \mathrm{e}^{-zx}\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x \\ &= \lambda \left[-\frac{\mathrm{e}^{-(z+\lambda)x}}{z+\lambda}\right]_0^\infty \\ &= \frac{\lambda}{z+\lambda}\,.\end{aligned}$$

Laplace Transform

- First-order derivative:

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathcal{L}_X(z) = \frac{\mathrm{d}}{\mathrm{d}z}\frac{\lambda}{z+\lambda} = -\frac{\lambda}{(z+\lambda)^2}\,.$$

- Thus,

$$\mathrm{E}[X] = \frac{\lambda}{(0+\lambda)^2} = \frac{1}{\lambda}\,.$$

- Second-order derivative:

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\mathcal{L}_X(z) = -\frac{\mathrm{d}}{\mathrm{d}z}\frac{\lambda}{(z+\lambda)^2} = \frac{2\lambda}{(z+\lambda)^3}\,.$$

- Thus,

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 = \frac{2\lambda}{(0+\lambda)^3} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}\,.$$

Laplace Transform

Lemma

Let X and Y be independent, nonnegative, continuous random variables. Then the Laplace transform of $W = X + Y$ is

$$\mathcal{L}_W(z) = \mathcal{L}_X(z)\,\mathcal{L}_Y(z)\,.$$

Proof:

$$\begin{aligned}\mathcal{L}_W(z) &= \mathrm{E}\left[\mathrm{e}^{-zW}\right] \\ &= \mathrm{E}\left[\mathrm{e}^{-zX}\,\mathrm{e}^{-zY}\right] \\ &= \mathrm{E}\left[\mathrm{e}^{-zX}\right]\mathrm{E}\left[\mathrm{e}^{-zY}\right] \\ &= \mathcal{L}_X(z)\,\mathcal{L}_Y(z) \qquad \square\end{aligned}$$

Moments and Laplace Transform

Summary

- Moments and central moments are quantities that characterize aspects of probability distributions.
- The Laplace transform is a mathematical tool that allows to relatively easily obtain moments of common continuous probability distributions.