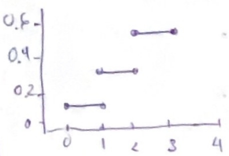
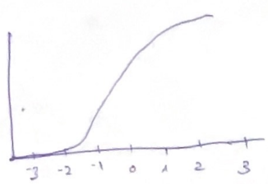
Unit 4: Continuous Random Variables

4.1 Probability density function



Discrete r.v



Continuous r.v

Definition 4.1.2 (continuous r.v)

- A r.v ~~is~~ has a continuous distribution if its CDF is differentiable (can get derivative at any point in CDF).
- Some endpoints where CDF is continuous but not differentiable are also allowed, as long as the CDF is differentiable everywhere else.
- A continuous r.v is a r.v with a continuous distribution

Definition 4.1.3 (~~PMF~~ PDF of continuous r.v)

- Consider r.v X with CDF F, the PDF of X is the derivative f of the CDF, given by $f(x) = F'(x)$.
- The support of X, and of its distribution, is the set of all x where $f(x) > 0$

Note:

- Key difference between discrete r.v and continuous r.v is that for a continuous r.v X, $P(X=x) = 0$ for all x. This is because $P(X=x)$ is the height of a jump at x, but the CDF of X has no ~~in~~ jumps
- Since CDF of continuous r.v has no jumps, its PMF would be 0, hence we work with PDF instead.
- Key difference between ~~&~~ PDF and PMF is for a PDF f, the quantity $f(x)$ is not a probability, and it is possible to have $f(x) > 1$. To obtain the probability, we need to integrate the PDF (calculate the area under the curve). Basically, getting from PDF back to CDF.

Proposition 4.1.4 PDF to CDF

Let X be a continuous r.v with PDF f. Then the CDF of X is given by

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

$$\text{PDF}(f(x)) \xrightarrow{\text{integrate } \int_{-\infty}^{x} f(x)\,dx = F(x)} \text{CDF}(F(x))$$

$$\text{CDF}(F(x)) \xrightarrow{\text{derivative } f(x) = \frac{d\,F(x)}{dx}} \text{PDF}(f(x))$$

PDF → determine the distribution

CDF → determine the probability

Note:

- We can be carefree about including or excluding endpoints for continuous r.v.

$$P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b)$$

- $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)\,dx$

Theorem 4.1.6 (Valid PDFs)

The PDF of a continuous r.v must satisfy these criteria:
- Non-negative: $f(x) \geq 0$
- Integrates to 1: $\int_{-\infty}^{\infty} f(x)\,dx = 1$

4.2 Uniform

Definition 4.2.1 Uniform distribution

A continuous r.v U is said to have Uniform distribution on interval (a, b) if its PDF is:

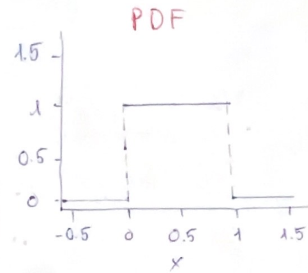
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

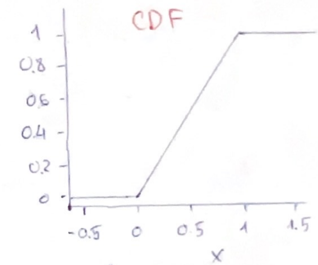
We denote this by $U \sim \text{Unif}(a, b)$.

The CDF is the accumulated area under the PDF:

$$F(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x < b \\ 1 & \text{if } x \geq b \end{cases}$$

The Standard Uniform distribution will be most frequently used, denoted as $\text{Unif}(0, 1)$. Its PDF and CDF are simple:
- $f(x) = 1$ for $0 < x < 1$
- $F(x) = x$ for $0 < x < 1$





$\text{Unif}(0,1)$: Standard Uniform

Proposition 4.2.3 (Subinterval)

Let $U \sim \text{Unif}(a,b)$ and let (c,d) be a subinterval of (a,b) of length l (so $l = d-c$). Then the probability of U being in (c,d) is proportional to l.

For example, a subinterval that is twice as long has twice the probability of containing U.

↳ Example: Given $U \sim \text{Unif}(0, 10)$ and subinterval $(1, 2)$, then subinterval $(1, 3)$ is twice the probability of containing U compare to $(1, 2)$.

Definition 4.2.4 Location-scale transformation

Let X be an r.v and $Y = \sigma X + \mu$, where σ and μ are constants with $\sigma > 0$. Then we say that Y is has been obtained as a location-scale transformation of X.

For example, if $X \sim \text{Unif}(1, 2)$, then $Y_1 = X + 5$ is Uniform on interval $(6, 7)$, $Y_2 = 2X$ is Uniform on interval $(6, 7)$

Warning 4.2.6 Apply Location-scale transformation

Location-scale should be applied on the random variable, not its PDF.

For example, given r.v $U \sim \text{Unif}(0,1)$ with PDF $f(x) = 1$ on $(0,1)$ and $f(x) = 0$ otherwise. Then $3U + 1 \sim \text{Unif}(1,4)$, but $3f + 1 = 4$ on $(0,1)$ and 1 elsewhere, which is not a valid PDF since it doesn't integrate to 1.

4.3 Universality of the Uniform

Theorem 4.3.1 (Universality of the Uniform)

Let F be a CDF which is a continuous function and strictly increasing on the support of the distribution. This ensures that the inverse function F^{-1} exists, as a function from $(0,1)$ to $\mathbb{R}$. We have the following results:

1. Let $U \sim \text{Unif}(0,1)$ and $X = F^{-1}(U)$. Then X is an r.v with CDF F.
2. Let X be an r.v with CDF F. Then $F(X) \sim \text{Unif}(0,1)$

Explaining the theorem:

- The first part said if we have a r.v $U \sim \text{Unif}(0,1)$ and a CDF F, then we can create a r.v X whose CDF is F by plugging U into the inverse function F^{-1}.
- The second part, if we have a r.v X whose CDF is F, then we can create a r.v with Standard Uniform distribution by plugging X into function F.

$$\begin{tikzcd} \text{r.v } X \arrow[r, bend right, "U = F(X)\ (\text{create } U \sim \text{Unif}(0,1))"'] & \text{r.v } U \sim \text{Unif}(0,1) \arrow[l, bend right, "X = F^{-1}(U)\ (\text{create r.v } X \text{ with CDF } F)"'] \end{tikzcd}$$

Example 4.3.3 (Percentiles)

A large number of students take on an exam graded on a scale from 0 to 100. Let X be the score of a random student. Suppose that X is continuous, with CDF F strictly increasing on $(0,100)$, and the median score is 60. Then:

- $F(60) = 1/2$ or $F^{-1}(1/2) = 60$

If Jimmy scores 72, then $F(72)$ falls somewhere between $(1/2, 1)$. Going the other way around, $F^{-1}(0.95)$ will give us the score that has that percentile.

Now we consider the act of plugging X to its own CDF (second point in theorem 4.3.1). $F(X)$ gives us the percentile attained by a student. If we consider the distribution of the percentiles, it is Uniform. In other word, $F(X) = U \sim \text{Unif}(0,1)$. For example, 50% ~~percen~~ of the students have a percentile of at least 0.5. Universality of the Uniform here means 10% of students have percentile between $(0, 0.1)$, $(0.1, 0.2)$, $(0.2, 0.3)$ and so on - basically stating the definition of percentile.

Example 4.3.4 (Universality with Logistic)

. The Logistic CDF is : $F(x) = \dfrac{e^x}{1+e^x}$, $x \in \mathbb{R}$

. Part 1 of the universality says that $F^{-1}(U) \sim$ Logistic with $U \sim \text{Unif}(0,1)$, so we first inverse CDF to get F^{-1}:

$$F^{-1}(u) = \log\left(\frac{u}{1-u}\right)$$

then plug U for u:

$$F^{-1}(U) = \log\left(\frac{U}{1-U}\right)$$

therefore : $F^{-1}(U) \sim$ Logistic

. Conversly, part 2 of the universality property says that if $X \sim$ Logistic, then :

$$F(X) = \frac{e^X}{1+e^X} \sim \text{Unif}(0,1)$$

4.4 Normal Distribution

A famous continuous distribution with bell-shaped PDF. It is extremely widely used in statistics because of the central limit theorem, which says that the sum of a large number of i.i.d random variables has approximately Normal distribution.

Definition 4.4.1 (Standard Normal Distribution)

A continuous r.v Z ~~is~~ has Standard Normal Distribution if its PDF φ is given by :

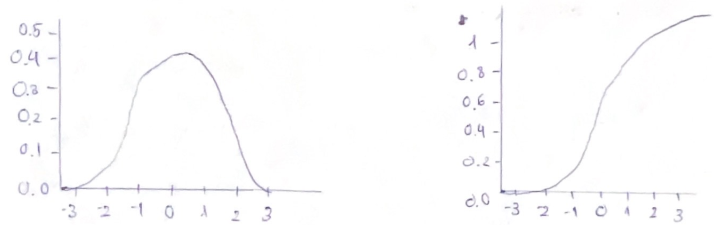
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} , \quad -\infty < z < \infty$$

Denoted as $Z \sim N(0,1)$

Note:

- The constant $\dfrac{1}{\sqrt{2\pi}}$ is needed to make the PDF integrate to 1. It is called "normalizing constants".
- The CDF Φ is the accumulated area under the PDF :

$$\Phi(z) = \int_{-\infty}^{z} \varphi(t)\, dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$



Standard Normal Distribution

Important symmetry properties :

- Symmetry of PDF : $\varphi(z) = \varphi(-z)$. φ is an even function.
- Symmetry of tail areas :
  The area under the PDF curve to the left of -2, which is $P(Z \le -2) = \Phi(-2)$, equals the area to the right of 2, which is $P(Z \geq 2) = 1 - \Phi(2)$. In general:
  $$\Phi(z) = 1 - \Phi(-z)$$
- Symmetry of Z and $-Z$:
  If $Z \sim N(0,1)$, then $-Z \sim N(0,1)$
  To prove this, the CDF of $-Z$ is:
  $P(-Z \le z) = P(Z \geq -z) = 1 - \Phi(-z) = \Phi(z) = P(Z \le z)$

The general Normal Distribution has 2 parameters, μ and σ^2 which are the mean and variance.

Start with a standard Normal r.v, ~~we can~~ $Z \sim N(0,1)$, we can convert to a Normal r.v with any desired params μ and σ^2 by location-scale transformation.

Definition 4.4.4 (Normal Distribution)

If $Z \sim N(0,1)$, then:

$$X = \mu + \sigma Z$$

is said to have Normal Distribution with $\begin{cases} \cdot \text{ mean } \mu \\ \cdot \text{ variance } \sigma^2 \end{cases}$

for $\mu \in \mathbb{R}$, $\sigma > 0$.

Denoted as $X \sim N(\mu, \sigma^2)$

We can also get from ~~a~~ non-standard Normal to standard Normal. In other words, get from X to Z, this is called "standardization".

Standardization

For $X \sim N(\mu, \sigma^2)$, the standardized version of X is:

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

We can use standardization to find the CDF and PDF of X in terms of standard Normal CDF and PDF

Theorem 4.4.5 (Normal CDF and PDF)

Let $X \sim N(\mu, \sigma^2)$. Then CDF of X is: $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

and the PDF of X is: $f(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$

Three important benchmarks for the Normal distribution are the probabilities of falling with one, two, and three standard deviations of the mean parameter μ.

The 68 - 95 - 99.7 rule tells us that these probabilities are what the name suggests.

Theorem 4.4.6 (68 - 95 - 99.7 rule)

If $X \sim N(\mu, \sigma^2)$, then:

$$P(|X-\mu| < \sigma) \approx 0.68$$

$$P(|X-\mu| < 2\sigma) \approx 0.95$$

$$P(|X-\mu| < 3\sigma) \approx 0.997$$

Example 4.4.7:

Let $X \sim N(-1, 4)$. What is $P(|X| < 3)$, exactly (in terms of Φ) and approximately?

- The event $|X| < 3$ is the same as $-3 < X < 3$
- ~~The~~ Using standardization, we can express this event in standard Normal r.v $Z = \left(\frac{X-(-1)}{2}\right)$
- The exact answer is:

$$P(-3 < X < 3) = P\left(\frac{-3-(-1)}{2} < \frac{X-(-1)}{2} < \frac{3-(-1)}{2}\right) = P(-1 < Z < 2)$$

$$= \Phi(2) - \Phi(-1) \approx 0.8186$$

- The approximate answer is:

By 68-95-99.7 rule, we know that: $\begin{cases} P(-1 < Z < 1) \approx 0.68 \\ P(-2 < Z < 2) \approx 0.95 \end{cases}$

$$\Rightarrow P(-1 < Z < 2) = P(-1 < Z < 1) + P(1 < Z < 2) = 0.815$$

Theorem 4.4.8 (Sum of Independent Normals)

If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, then:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

4.5 Exponential Distribution

The Exponential distribution is a simple model used for the waiting time for a certain kind of event to occur, e.g., the time until the next email arrives

Definition 4.5.1 (Exponential Distribution)

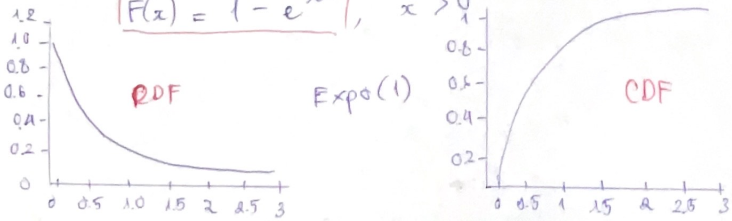
A continuous r.v X is said to have the Exponential distribution with parameter λ $(\lambda > 0)$ if its PDF is:

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

Denoted as $X \sim \text{Expo}(\lambda)$

The ~~core~~ ~~F~~ corresponding CDF is:

$$\boxed{F(x) = 1 - e^{-\lambda x}}, \quad x > 0$$



Exponential Distribution

Note:

- Exponential r.v cannot be shifted since it is defined to have support $(0, \infty)$ and shifting would change the left point.

- Exponential r.v can be scaled, we can use scalling to get from simple Expo(1) to general Expo(λ).
  - If $X \sim \text{Expo}(1)$, then $Y = \frac{X}{\lambda} \sim \text{Expo}(\lambda)$
  - Conversely, if $Y \sim \text{Expo}(\lambda)$, then $\lambda Y \sim \text{Expo}(1)$
- Exponential distribution has a property called memoryless, it says that if the waiting time for a certain event to occur is Exponential, no matter how long you have waited, the time remaining is still Exponential (with same parameter)

Definition 4.5.3 (Memoryless property)

A distribution is said to have memoryless property if a r.v X from that distribution satisfy:

$$P(X \geq s + t \mid X \geq s) = P(X \geq t)$$

for all $s, t > 0$

Example of memoryles property:

The wait time until the next train arrives has Exponential distribution ~~the~~ then condition on you having waited 30 minutes, the ~~b~~ train isn't due to arrive soon. The distribution simply forgets that you have been waiting for 30 minutes, and your ~~remaining~~ remaining wait time is the same as if you have just arrived at the station.

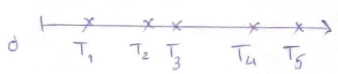
4.6 Poisson processes

The Exponential and Poisson distributions are closely related, as suggested by the use of λ. They are linked by a common story, the Poisson process.

Definition 4.6.1 (Poisson process)

A process of arrivals in continuous time is called a Poisson process with rate λ if the following 2 conditions hold:

- The number of arrivals that occur in an interval t is a $Pois(\lambda t)$ random variable.
- The number of arrivals that occur in disjoint intervals are independent of each other. For example, the number of arrivals in intervals $(0, 10)$, $[10, 15)$ and $[15, \infty)$ are independent.



Poisson process

Consider example of the arrivals of emails landing in an inbox. Suppose this follows the Poisson process with rate λ, we can ask these questions:

- How many emails arrive in one hour?

The answer comes directly from the definition, which tells us that the number of emails arrived in an hour follows a $Pois(\lambda)$ distribution. Number of email is nonnegative integer, so a discrete distribution is appropriate.

- How long does it take until the first email arrive?

The waiting time is a positive real number, so a continuous distribution on $(0, \infty)$ is ~~appriate~~ appropriate.

To find the distribution of T_1 (first email arrive time), we need to understand 1 crucial fact: saying arrival time of first email greater than t ($T_1 > t$) is equals no email arrived between 0 and t. So:

$$T_1 > t \text{ is same event as } N_t = 0$$

(N_t is number of email arrived in $(0, t]$)

This is called count-time duality because it connects a discrete r.v N_t (which counts the number of arrivals) with a continuous r.v T_1 (which marks the time of first arrival).

Since 2 events are the same, they have the same probability, since $N_t \sim Pois(\lambda t)$ by definition of Poisson process:

$$\therefore P(T_1 > t) = P(N_t = 0) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

$$\Rightarrow P(T_1 \leq t) = 1 - e^{-\lambda t}$$

So $T_1 \sim Expo(\lambda)$. The time until the first arrival in a Poisson process of rate λ has an Exponential distribution with parameter λ.

• What about $T_2 - T_1$, the time between the first and second arrivals?

• Since disjoint intervals in a Poisson process are independent by definition, the past is irrelevant once the first arrival occurs, thus $T_2 - T_1 \sim \text{Expo}(\lambda)$, its independent of the time until the first arrival.

• Similarly, $T_3 - T_2 \sim \text{Expo}(\lambda)$ independently of T_1 and $T_2 - T_1$

• Continue this logic, we can deduce that all interarrival times are i.i.d $\text{Expo}(\lambda)$ random variables.

Summarize:

In a Poisson process of rate λ:

• The number of arrivals in interval of length 1 is $\text{Pois}(\lambda)$

• The times between arrivals are i.i.d $\text{Expo}(\lambda)$

Example 4.6.3 (Minimum of independent Expos)

Let $X_1, \dots, X_n$ be independent, with $X_j \sim \text{Expo}(\lambda_j)$. Let $L = \min(X_1, \dots, X_n)$. Show that $L \sim \text{Expo}(\lambda_1 + \dots + \lambda_n)$, and interpret this intuitively.

Solution:

Consider its survival function $P(L > t)$, since the survival function equals $1 - \text{CDF}$.

$$P(L > t) = P(\min(X_1, \dots, X_n) > t)$$
$$= P(X_1 > t, \dots, X_n > t)$$
$$= P(X_1 > t) \dots P(X_n > t)$$
$$= e^{-\lambda_1 t} \dots e^{-\lambda_n t}$$
$$= e^{-(\lambda_1 + \dots + \lambda_n)t}$$

Intuitively, we can intepret λ_j as the rates of n independent Poisson processes. We can imagine, for example, X_1 is the waiting time for a green car, X_2 is the waiting time for a blue car and so on. Then L is the waiting time for a car of any ~~of the~~ colors to pass by, so it makes sense that L has a combined rate of $\lambda_1 + \dots + \lambda_n$

In other words, each exponential distribution X_i represents an independent event occuring at a rate λ_i. When multiple independent events are possible, the minimum waiting time L to the first event is still exponential, but with a rate equals to the sum of individual rates. This is because the rates add up, ~~so~~ meaning overall likelihood of any event happening increases accordingly.