Important knowledge

This is the mathematical and probability knowledge required for this course:

The mathematics part should have been covered in pre-requisite courses. Coverage of probability is less certain (pun intended 😉) and needed in the second half of the course, so you should review the `second part below` with particular care.

1. Revisions Part 1: Mathematics

1.1. Functions and their derivatives

1. Be familiar with functions $x^\alpha, e^{\alpha x}, \ln(1+x)$
2. Basic derivatives:

$$(x^\alpha)' = \alpha x^{\alpha-1}$$

$$(e^{\alpha x})' = \alpha e^{\alpha x}$$

$$(\ln(1+x))' = \frac{1}{1+x}$$

$$(a^x)' = a^x \ln(a)$$

3. `Taylor's expansion` (here for an exponential random variable):

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + \ldots$$

1.2. Be able to find expressions for following summations

$$\sum_{i=1}^{n} x^i, \sum_{i=1}^{n} ix^i, \sum_{i=1}^{n} i$$

See Tutorial 0 (Revisions) for solutions.

1.3. Change the order of double summation

$$\sum_{k=1}^{n}\sum_{j=1}^{k} a_{k,j} = \sum_{j=1}^{n}\sum_{k=j}^{n} a_{k,j}$$

1.4. The solution to a quadratic equation

The equation

$$ax^2 + bx + c = 0$$

has two solutions:

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \qquad b^2 > 4ac$$
$$x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \qquad b^2 > 4ac$$

If $b^2 - 4ac$ = 0, the equation has a double solution: $x = \frac{-b}{2a}$

1.4.1. Example

For example, find a value for v such that $v \in (0, 1)$ and v satisfies the equation: $v - 2v^{0.5} + \dfrac{3}{4} = 0$.

Solution: let $v^{\frac{1}{2}} = x$. Then the equation above simplifies to $x^2 - 2x + \frac{3}{4} = 0$ which has two solutions:

$$x_1 = \frac{2 + \sqrt{4 - 3}}{2} = 1.5, \qquad x_2 = \frac{2 - \sqrt{4 - 3}}{2} = 0.5$$

We reject the solution $x_1 = 1.5$ as $x_1 > 1$. Then $v = x_2^2 = 0.5^2 = 0.25$ is the required solution.

1.5. Be able to solve simple differential equations

1.5.1. Example 1

For example, solve

$$f'(x) = 2x$$

with initial condition $f(0) = 1$.

Solution:

$$\begin{aligned} f(x) =& f(0) + \int_0^x f'(t)dt \\ =& 1 + \int_0^x 2tdt \\ =& 1 + t^2|_0^x \\ =& 1 + x^2 \end{aligned}$$

1.5.2. Example 2

Solve

$$f'(x) = 2f(x)$$

with initial condition $f(0) = 1$.

Solution:

$$\frac{f'(x)}{f(x)} = 2 \Longleftrightarrow (\ln f(x))' = 2$$

.

$$\Rightarrow \ln f(x) - \ln f(0) = \int_0^x (\ln f(t))' dt = \int_0^x 2dt = 2x$$

$$\Rightarrow \ln f(x) = 2x \Rightarrow f(x) = e^{2x}$$

1.6. Integrals

1. We have

$$\int_a^b f(x)dx = F(b) - F(a),$$

   where $F(x)$ is the anti-derivative of $f(x)$, such that $F'(x) = f(x)$.
2. The integration variable is just a tool, that is,

$$\int_a^b f(x)dx = \int_a^b f(y)dy,$$

   it does not matter to use x or y.
3. We have

$$\$ \int_0^\infty f(x)dx = \lim_{b\to\infty} \int_0^b f(x)dx = \lim_{b\to\infty} F(b) - F(0).$$

   For example:

$$\begin{aligned}\int_0^\infty e^{-x}dx =& \lim_{b\to\infty}\int_0^b e^{-x}dx \\ =& \lim_{b\to\infty}\left(-e^{-x}|_0^b\right) \\ =& 1 - \lim_{b\to a} e^{-b} \\ =& 1\end{aligned}$$

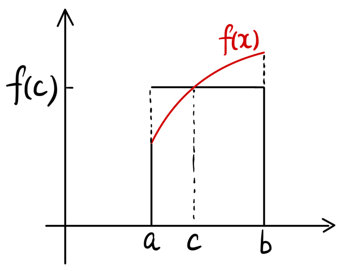
4. `Integration by parts`:

$$\begin{aligned}\int_a^b f(x)g'(x)dx =& f(x)g(x)|_a^b - \int_a^b g(x)f'(x)dx \\ =& f(b)g(b) - f(a)g(a) - \int_a^b g(x)f'(x)dx\end{aligned}$$

1.7. The average of a function on $[a, b]$

Let $f(x)$ be a continuous function on $[a, b]$. Then there exists a point $c \in [a, b]$ such that

$$\int_a^b f(x)dx = f(c)(b-a), \qquad c \in [a, b]$$

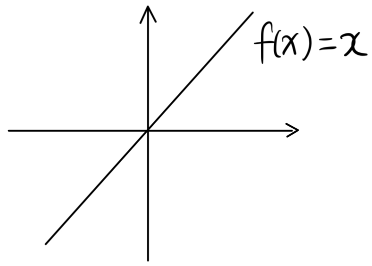


Interpretation: $\int_a^b f(x)dx$ is the area of the region enclosed by $f(x)$, x-axis, $x = a$, $x = b$. $f(c)(b-a)$ is the area of the rectangle of height $f(c)$ and length $(b-a)$.

Definition: $\frac{\int_a^b f(x)dx}{b-a} = f(c)$: the average value of $f(x)$ on $[a, b]$ interval.

1.7.1. Example 1:

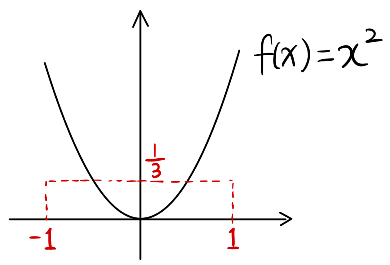
$$\frac{\int_{-1}^{1} x dx}{2} = 0$$



The average value of $f(x) = x$ on $[-1, 1]$ is 0.

1.7.2. Example 2:

$$\frac{\int_0^1 x dx}{1-0} = \frac{1}{2}$$



The average value of $f(x) = x$ on $[0, 1]$ is $\frac{1}{2}$.

1.7.3. Example 3:

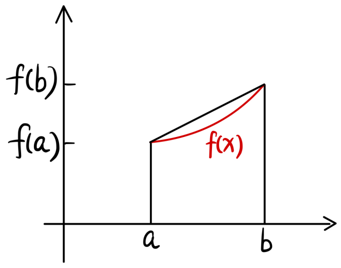
$$\frac{\int_{-1}^{1} x^2 dx}{2} = \frac{1}{3}$$

The average value of $f(x) = x^2$ on $[-1, 1]$ is $\frac{1}{3}$.

1.8. The trapezoid rule in integration

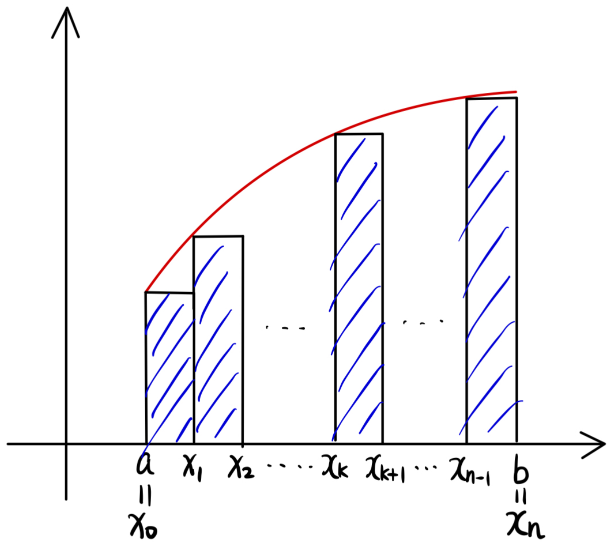
$$\int_a^b f(x)dx \approx \frac{1}{2}[f(b)+f(a)](b-a)$$

$$\iff \frac{\int_a^b f(x)dx}{b-a} \approx \frac{1}{2}[f(b)+f(a)]$$



The average value of $f(x)$ on $[a,b]$ can be approximated by $\frac{1}{2}[f(b)+f(a)]$.

1.9. The definition of $\int_a^b f(x)dx$ and its numerical calculations



$$\int_a^b f(x)dx = \lim_{n\to\infty}\sum_{k=0}^{n-1} f(x_k)\frac{b-a}{n}, \qquad (1)$$

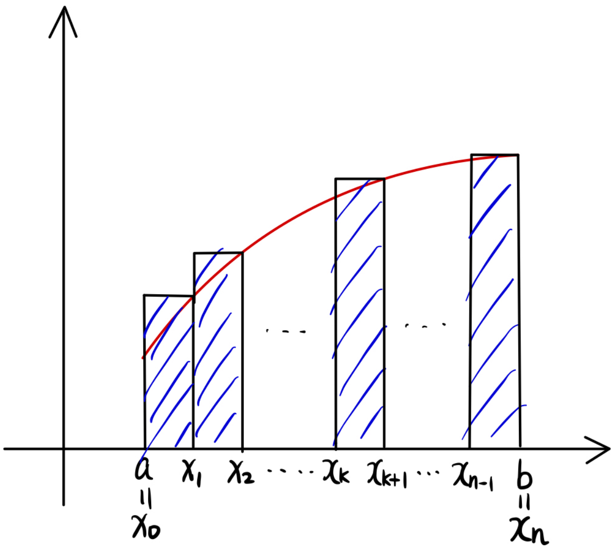
where $x_0 = a, x_1 = x_0 + \frac{b-a}{n}, \dots, x_{k+1} = x_k + \frac{b-a}{n}, \dots, x_n = b$

In the summation in (1), each term represents the area of a rectangle. $f(x_k)\frac{b-a}{n}$ represents the area of the k-th rectangle.

Approximations:

1. $\int_a^b f(x)dx \approx f(a)(b-a) \qquad (n=1)$
2. $\int_a^b f(x)dx \approx \frac{b-a}{2}\left[f(a)+f\left(\frac{b+a}{2}\right)\right]$
3. $\int_a^b f(x)dx \approx (b-a)\frac{f(x_0)+f(x_1)+\ldots+f(x_{n-1})}{n}$: the average of $f(x_0), f(x_1), \dots, f(x_{n-1})$ times $(b-a)$.
4. If $a=0, b=1$, $\int_0^1 f(x)dx \approx \frac{f(x_0)+f(x_1)+\ldots+f(x_{n-1})}{n}$.

Alternatively,



$$\int_a^b f(x)dx = \lim_{n\to\infty}\sum_{k=1}^{n} f(x_k)\frac{b-a}{n} \qquad (2),$$

where $x_1 = x_0 + \frac{b-a}{n}, x_2 = x_1 + \frac{b-a}{n}, \ldots, x_n = b$.

Approximations:

1. $\int_a^b f(x)dx \approx f(b)(b-a)$
2. $\int_a^b f(x)dx \approx \frac{b-a}{2}\left[f\left(\frac{b+a}{2}\right) + f(b)\right]$
3. $\int_a^b f(x)dx \approx (b-a)\frac{f(x_1)+f(x_2)+\ldots+f(x_n)}{n}$

1.10. The average number of n numbers

Let $x_1, x_2, \ldots, x_n$ be n numbers. Then

$$\frac{x_1+\ldots+x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is the average value of $x_1, x_2, \ldots, x_n$.

1.10.1. Example 1

The average value of $1, 2, \ldots, n$ is

$$\frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2},$$

where $1+2+3+\ldots+n = \frac{n(n+1)}{2}$ is given in Tutorial 0.

1.10.2. Example 2

One student took 8 subjects in his first year at University of Melbourne. The results are as follows: Semester 1: 75, 83, 65, 90; Semester 2: 60, 76, 80, 50.

Then

- $75+83+65+90+60+76+80+50 = 579$ is the total marks from year 1.

- The average mark is $\frac{579}{8} = 72.4$
- The average mark for S1 is $\frac{75+83+65+90}{4} = 78.2$
- The average mark for S2 is $\frac{60+76+80+50}{4} = 66.5$

1.11. The weighted average of n numbers

Let $x_1, x_2, \ldots, x_n$ be n real numbers.

Let $\theta_1, \theta_2, \ldots., \theta_n$ be n numbers such that

$$0 \leq \theta_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{n} \theta_i = 1.$$

Then

$$\sum_{i=1}^{n} \theta_i x_i$$

is called the *weighted average* of $x_1, x_2, \ldots, x_n$.

Note:

- θ_i is the weight attached to x_i.
- if $\theta_i = \frac{1}{n}$, then

$$\sum_{i=1}^{n} \frac{1}{n} x_i = \frac{1}{n} \sum_{i=1}^{n} x_i$$

  is the average of $x_1, x_2, \ldots, x_n$ (equally weighted).

1.11.1. Example

In the assessment of ACTL10001, the assignments account for 20%, the mid-semester exam accounts for 10%, and the final exam accounts for 70%. A student got 70 out of 100 for mid-semester result, 95 out of 100 for assignments, and 80 for final exam. Then the overall weighted average mark is

$$70 \times 10\% + 95 \times 20\% + 80 \times 70\% = 82.$$

2. Revisions Part 2: Probability

2.1. Events and Probability

2.1.1. Vocabulary: events vs probability

It is important to understand the difference between events and probability:

- Event: what could happen - an actual “thing”, in real life, that could happen;
- Probability: our understanding of the “likelihood” (or frequency) of an event (something that could occur).

So when we are building a mathematical model for uncertain outcomes:

1. The first step is to work out what are all the possible things that could occur (for instance, “rain” or “no rain”). The full set of those is denoted Ω.
2. The second step is to make assumptions about how likely those things can occur. Here "Pr" is an operator that maps an event into a probability. For instance, $\mathrm{Pr}[\mathrm{rain}] = 0.2$ means that the likelihood corresponding to the event “rain” is 20%.

In what follows we outline basic results and axioms around events and their probabilities. Often logic means that a result or definition on one side (e.g. events) can be translated on the other side (e.g. probabilities).

For instance, the complement to an event is exactly whatever could happen, that is not the event. Hence, the probability of the complement must be 1 minus the probability of the original event; see 2.1.2.4 below.

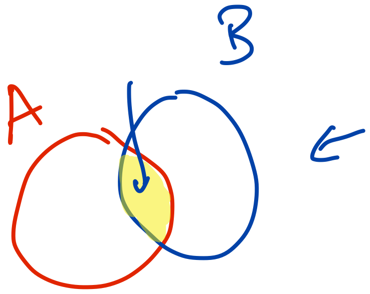
2.1.2. Events, operations of events, probability of an event

1. $\emptyset$: empty set, that it, it is an impossible event:

$$\mathrm{Pr}(\emptyset) = 0.$$

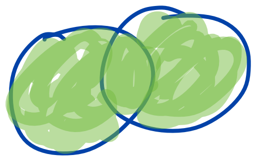
2. Ω: the full set of possible outcomes, that is, it is a certain event:

$$\mathrm{Pr}(\Omega) = 1.$$



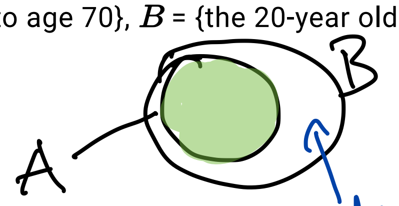
3. A: an event (within Ω), $0 \leq \mathrm{Pr}(A) \leq 1$.
4. A^C: the event that A does not occur (called a “complement”):

$$\mathrm{Pr}(A^C) = 1 - \mathrm{Pr}(A).$$



5. $A \cap B$: A **and** B, the event that both A and B occur.
6. $A \cup B$: A **or** B, the event that either A or B, or both events occur.
7. $A \subseteq B$: If A occurs, B must, and:
   - $\mathrm{Pr}(A) \leq \mathrm{Pr}(B)$;
   - $A \cap B = A$.

   Example: A = {a 20-year old survives to age 70}, B = {the 20-year old survives to age 50}. Then $A \subseteq B$.



dying between 50 and 70

2.1.3. Mutually exclusive events

If

$$A \cap B = \emptyset,$$

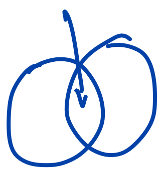
then A and B are **mutually exclusive**. Also,

$$\mathrm{Pr}(A \cup B) = \mathrm{Pr}(A) + \mathrm{Pr}(B).$$

2.1.4. Independent events A and B

If A and B are independent, then

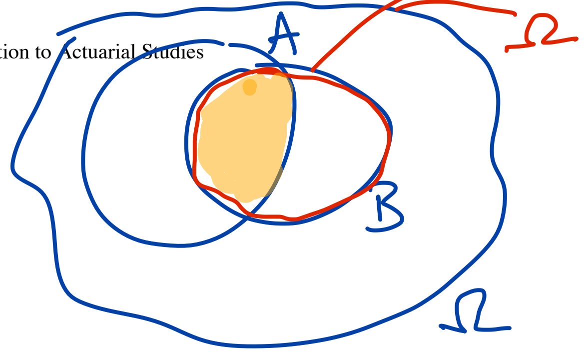
$$\mathrm{Pr}(A \cap B) = \mathrm{Pr}(A)\,\mathrm{Pr}(B).$$



2.1.5. Conditional probability formula

We have

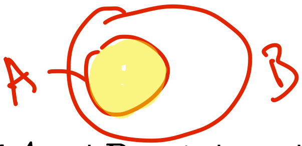
$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$



This leads to Bayes' theorem, see for instance `this`.

Also,

1. If $A \subseteq B$, then $A \cap B = A$ and

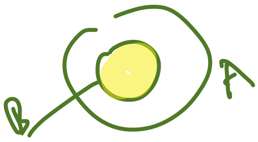


$$\Pr(A|B) = \frac{\Pr(A)}{\Pr(B)}.$$

2. If A and B are independent, then

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)\Pr(B)}{\Pr(B)} = \Pr(A).$$

3. If $B \subseteq A$, then $A \cap B = B$ and



$$\Pr(A|B) = \frac{\Pr(B)}{\Pr(B)} = 1.$$

Given B has occurred, A is certain.

2.2. Random variables and their distribution

"Pr"

2.2.1. Definition

A random variable, denoted by capital letters X, Y, Z, is a quantity whose value is subject to variations due to chance.

2.2.2. Distribution Function

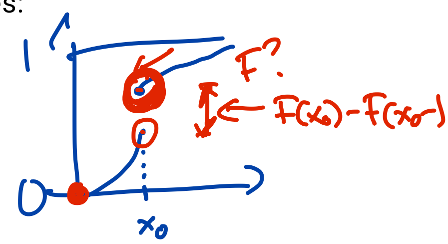
Definition:

$$F(x) = \Pr(X \leq x) \quad x \in \mathbb{R}$$

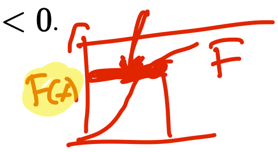
Ω

$F(x)$ is called the distribution function of X, and it has the following properties:

1. $F(-\infty) = 0, F(\infty) = 1.$
2. $F(x_1) \leq F(x_2)$, if $x_1 \leq x_2$.
3. $F(x)$ is right-continuous (aka "càdlàg"), i.e., $\lim_{x \to x_0^+} F(x) = F(x_0)$.
4. $F(b) - F(a) = \Pr(a < X \leq b)$. $= \Pr[X \leq b] - \Pr[X \leq a]$
5. $F(b-) - F(a) = \Pr(a < X < b)$.
6. $F(b) - F(a-) = \Pr(a \leq X \leq b)$.
7. $F(b) - F(b-) = \Pr(X = b) \geq 0$



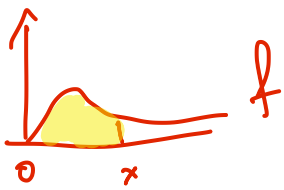
In our subject, we generally assume that $X \geq 0$ so that $F(x) = 0$ for $x < 0$.



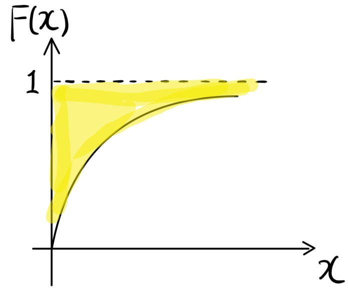
2.2.3. Continuous random variables ($X \geq 0$)

X is said to be a continuous r.v. if X has a probability density function $f(x)$, $x \geq 0$, with the following properties:

1. $f(x) = F'(x)$.



2. $F(x) = \int_0^x f(y)dy$.
3. $\Pr(a < X \le b) = \Pr(a < X < b) = \Pr(a \le X \le b) = \Pr(a \le X < b) = \int_a^b f(x)dx$ $= F(b) - F(a)$
4. $F(x)$ typically looks like



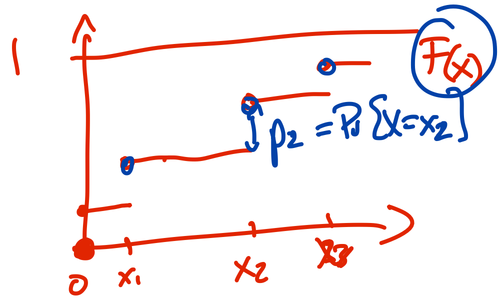
but note that it does not need to be concave.

5. $E(X) = \int_0^\infty x f(x)dx = \int_0^\infty [1 - F(x)]dx$

$E? [f(\text{event})]$
$= \sum_{\text{all events in } \Omega} f(\text{"event"}) \Pr(\text{"event"})$

2.2.4. Discrete random variables

A random variable X is said to be a discrete random variable if X takes values from a countable set of numbers $\{x_1, x_2, \dots, x_n, \dots\}$.

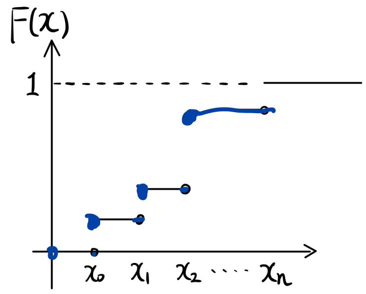


1. Probability distribution of X

X	x_1	x_2	...	x_n	...
prob	p_1	p_2	...	p_n	...

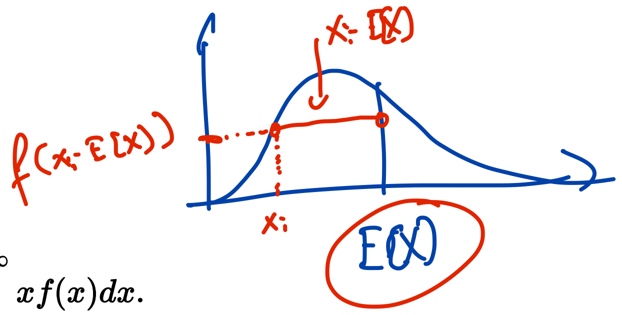
where $p_n = \Pr(X = x_n), n = 1, 2, 3, \dots$

2. $E(X) = \sum_{n=1}^\infty x_n p_n$
3. The distribution function $F(x)$ is a piece-wise constant function (also called step function).



2.3. Moments of a random variable

2.3.1. Expectation and variance



Expectation of X:

$$E(X) = \int_0^\infty x f(x)dx.$$

Variance of X:

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

Furthermore:

1. $Var(X)$ measures the variability of X. The larger the variance, the more variability X has.

2. If $Var(X) = 0$, $X \equiv c \equiv E(X)$. There is no variability for X. X is a constant.
3. If X and Y are independent, then

$$Var(X + Y) = Var(X) + Var(Y).$$

4. $Var(aX) = a^2 Var(X)$

$$Var(aX) = E[(aX - E[aX])^2] = E[(aX - aE[X])^2]$$
$$= E[a^2(X - E[X])^2] = a^2 E[(X - E[X])^2]$$
$$= a^2 Var(X)$$

2.3.2. Moments of the average of iid rv's

Assume $X_1, X_2, \ldots, X_n$ are independently and identically distributed, with

$$E(X_1) = \mu, \text{ and } Var(X_1) = \sigma^2.$$

Define

$$Y_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

to be the average of $X_1, X_2, \ldots, X_n$. Then

$$E(Y_n) = \frac{1}{n}(E(X_1) + E(X_2) + \ldots + E(X_n)) = \frac{1}{n}(\mu + \mu + \ldots + \mu) = \mu$$

and

$$= Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} Var\left(\sum X\right)$$

n times

$$Var(Y_n) = \frac{1}{n^2}(Var(X_1) + Var(X_2) + \ldots + Var(X_n)) = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \ldots + \sigma^2) = \frac{\sigma^2}{n}.$$

When $n \to \infty$, $Var(Y_n) \to 0$. That is to say, as $n \to \infty$, $Y_n \to \mu$. With an infinite sample of X's, you can estimate μ with certainty.

2.4. Selected distributions

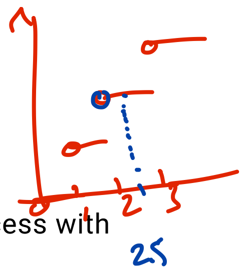
2.4.1. Binomial distribution

If $X \sim Bin(n, p)$, then

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 < p < 1, k = 0, 1, 2, \ldots, n.$$

Note:

1. $E(X) = np$, $Var(X) = np(1-p)$.
2. $F(k) = \Pr(X \le k) = \sum_{j=0}^{k} \Pr(X = j) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j}, \quad k = 0, 1, 2, \ldots, n.$
3. $F(2.5) = \Pr(X \le 2.5) = \Pr(X \le 2) = F(2)$.
4. X represents # of successes out of n independent trials, each trail has two outcomes: sucess with probability p OR failure with probability $1 - p$.



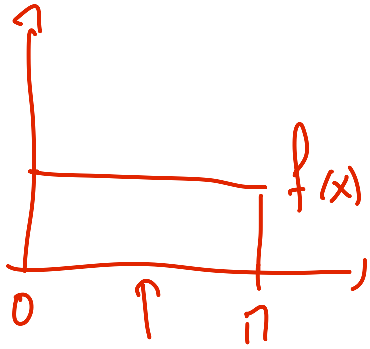
2.4.2. Exponential distribution

If $X \sim Exp(\lambda)$ then

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

Note:

1. $F(x) = 1 - e^{-\lambda x}, x \geq 0$.
2. $E(X) = \int_0^\infty x f(x) dx = \int_0^\infty [1 - F(x)] dx = \frac{1}{\lambda}$.

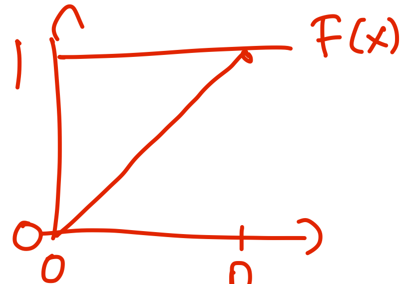


2.4.3. Uniform distribution

If $X \sim U(0, M)$ then

$$f(x) = \frac{1}{M}, 0 \leq x \leq M.$$

Note:



1. $F(x) = \frac{x}{M}, 0 \leq x \leq M$.
2. $E(X) = \int_0^M x \frac{1}{M} dx = \int_0^M (1 - \frac{x}{M}) dx = \frac{M}{2}$.

3. Credit

The initial version of those notes were developed by Professor Shuanming Li in 2018. These were then transcribed modified and augmented by Professor Benjamin Avanzi in 2021.