# Applied Statistics 10-09-2019

Let A,B,C be finite sets.
**Objective:** Count the union of A,B,C, therefore $\#(AUBUC)$
$\#(AUBUC)=\#(AU(BUC))$
$=\#(A) + \#(B\cup C) - \#(A\cap(B\cup C))$
$=\#(A) + \#(B)+\#(C) - \#(B\cap C) - \#((A\cap B)\cup(A\cap C))$
$=\#(A) + \#(B)+\#(C) - \#(B\cap C) - \#(A\cap B) - \#(A\cap C) + \#((A\cap B)\cap(A\cap C))$
$=\#(A) + \#(B) + \#(C) - \#(B\cap C) - \#(A\cap B) - \#(C\cap A) + \#((A\cap A)\cap B\cap C)$

**Thm:** (Principle of inclusion-exclusion): For each finite set A,B,C
$\#(A\cup B\cup C)=\#(A) + \#(B) + \#(C)$` `- $(A\cap B) - \#(B\cap C) - \#(C\cap A)$` `$+ \#(A\cap B\cap C)$

**Thm:** (Principle of inclusion-exclusion):
$\#(A\cup B\cup C\cup D)=$
$\#(A) + \#(B) + \#(C) +\#(D) - \#(A\cap B) - \#(A\cap C)- \#(A\cap D) - \#(B\cap C) - \#(B\cap D) - \#(C\cap D) +$
$\#(B\cap C\cap D) + \#(A\cap C\cap D) + \#(A\cap B\cap D) + \#(A\cap B\cap C) - \#(A\cap B\cap C\cap D)$

## Principle of multiplication

Thm: (Principle of multiplication): For each finite set A,B:
$\#(A\times B)=\#(A)\cdot\#(B)$

**Proof:** We prove by induction on #(B).
Step 1: To show the formula when #(B)=0, shows that the proof is trivial.
Step 2: We assume that $\#(A\times B)=\#(A)\cdot\#(B)$, whenever #(B)=k. We plan to show this formula when $\#(B')=k+1$
**Proof:** Write $B'=\{b\}\cup B$ for some b $\in B'$ and b $\notin B$, and $\#(B)=k$.
Now,
$\#(A\times B')=\#(A\times(\{b\}\cup B))$
$=\#(A\times\{b\}\cup A\times B)$
$=\#(A\times\{b\}) + \#(A\times B)$
$=\#(A) + \#(A) + \#(B)$
$=\#(A)(\#(B)+1)$
$=\#(A)\#(B')$ The formula is valid when $\#(B')=k+1$.

By use of step 1 it can be seen that the above holds for $\#(B)=0$, and if that holds then by recursion and relation between $B$ and $B'$ the proof holds for all $\#(B)$.
So by mathematical induction $\#(A\times B)=\#(A)\cdot\#(B)$.

**Remark**: To create a member in $A\times B$,
Step 1: Choose a member $a\in A$ (There are $\#(A)$ amount of choices)
Step 2: Choose a member $b\in B$ (There are $\#(B)$ amount of choices)
Then $(a,b)\in A\times B$ is created and there are $\#(A\times B)$.

**Eg:** Compute the number of straight flshes when 5 cards are picked from the standard poker deck. (Amount of cards $4\cdot 13=52$)
**Soln:** To create a straight flush,

Step 1: Choose 1 among each suit. (There are 4 choices then.)
Step 2: Choose a face value among A,2,3...10. (There are 10 choices)
So, a straight flush is obtained by having 5 cards with common suit in step 1 and 5 consecutive face values initiated by one of the cards of step 2.
By the principle of multiplication there are $4 \cdot 10 = 40$ straight flushes.

**Eg:** Compute the number of 4 of a kind when 5 cards are picked.
**Soln:** To create 4 of a kind,
Step 1: Pick a face value among A,2,3...J,Q,K. (There are 13 choices)
Step 2: Pick a card whose face value is different from the one from step 1. (There are 48 choices)

By the principle of multiplication, there are $13 \cdot 48 = 624,$ 4 of a kind

**Def:** Let A,B be sets. A function f from A to B.
(symbolically, it is written as $f:(A \to B)$ is a rule of assignment, which assigns a member in B to each of the members in A.
**Eg:** Compute the number of functions from A to B.
**Soln:** Write $A = \{a_1, a_2 ... a_n\}$.
Step 1: Assign a member in B to $a_1$. (There are #(B) choices)
Step 2: Assign a member in B to $a_2$. (There are #(B) choices).
And so on...
By the principle of multiplication, there are $\#(B)^{\#(A)}$ functions from A to B.

## Sample space

**Def:** A sample soace is the set of all possible outcomes when an experiment is conducted.

**Eg:** Whena a coin is tossed. The sample space of this experiment is {H,T}.

**Eg:** When a die is rolled. The sample space of this experiment is {1,2,3,4,5,6}.

**Eg:** An American is picked randomly. The sample space is every American.

**Eg:** The waiting time of the bus 91M is recorded when one leaves today. The sample space is $|\Re|$, any real positive number.

**Def:** In an experiment, an event is a subset of it's sample space.

**Eg:** A die is rolled. It's sample space is {1,2,3,4,5,6}. Where {2,4,6} is an event, it is also described as the event of obtaining an even number.
{1,3,5} is also an event, which is described as an event of obtaining an odd number.

**Eg:** A die is rolled twice.Then the sample space is {(1,1),(1,2),(1,3),(1,4)...(6,6)} = {1,2,3,4,5,6}× {1,2,3,4,5,6}. Cartesian product obtains pairs of the sets.
Also written as $\{1, 2, 3, 4, 5, 6\}^2$

# Applied statistic 12-09-2019

## probabilities

**Eg:** In an experiment, a die is rolled and a coin is flipped. Its sample space = {(1,H),(1,T),(1, H),(2,T),...} = {1,2,...,6}×{H,T}
The event of obtaining an even number and a head is written as:
$=\{(2,H),(4,H),(6,H)\}$ or $=[2,4,6\}\times\{H\}$

**Eg:** In an experiment, the number of people attending a lecture is recorded. Its sample space is then:
$=\mathbb{N}\cup\{0\}$
$\mathbb{N}$ is the event, in which some people come (All whole numbers).

**Def:** Let S be the sample space of an experiment. Probability "$P$" is a function, from the set of all events to [0,1] satisfying the axioms:
Axiom 1) $P(S)=1$ and $P(\varnothing)=0$
Axion 2) If $A_1, A_2 ...$ are events such that $A_i \cap A_j = \varnothing$ for all $i \neq j$, Then
$P\left(A_1 \cup A_2 \cup ...\right) = P\left(A_1\right) + P\left(A_2\right) + ...$
**Remark:** The events $A_1, A_2 ...$ are mutually exclusive if $A_i \cap A_j = \varnothing$ for all $i \neq j$. This is the condition in Axion 2).
**Lemma:** If S is a sample space and AcBcS. Then P(A)≤P(B).
**Proof:** Observe that A, B\A are mutually exclusive, or disjoint, so that
$P(A\cup(B\backslash A)) = P(A) + P(B\backslash A) \geq P(A)$.
**Lemma:** $P\left(A^c\right) = 1 - P(A)$
**Proof:** Observe that $A$, $A^c$ are mutually exclusive. So $P\left(A\cup A^c\right) = P(A) + P\left(A^c\right) = P(S) = 1$.

**Lemma:** (Principle of Inclusion-exclusion) If A,B are events,
$P(A\cup B) = P(A) + P(B) - P(A\cap B)$
**Proof:** Observe that A,B\A are mutually exclusive, So
$P(A\cup B) = P(A\cup(B\backslash A)) = P(A) + P(B\backslash A)$
Observe also that B\A, A∩B are mutually exclusive. So
$P((B\backslash A)\cup(A\cap B)) = P(B\backslash A) + P(A\cap B)$
So that $P(B\backslash A) = P(B) - P(A\cap B)$.

**Def:** Let S be a finite sample space. P is the uniform probability if $P(\{a\}) = P(\{b\})$ for all $a, b \in S$.
Let $S=\left\{a_1, a_2 ... a_n\right\}$ $a_i \neq a_j$ for $i \neq j$.
$P(S) = 1 = P\left(\left\{a_1\right\}\cup\left\{a_2\right\}...\left\{a_n\right\}\right) = P\left(\left\{a_1\right\}\right) + P\left(\left\{a_2\right\}\right) + ... + P\left(\left\{a_n\right\}\right) = n\cdot P\left(\left\{a_1\right\}\right)$
So $P\left(\left\{a_1\right\}\right) = \frac{1}{n} = \frac{1}{\#`(`S)}$
If $AcS$ so that $A=\left\{b_1, ..., b_k\right\}$

$P(A) = P(\{b_1\} \cup ... \cup \{b_k\}) = P(\{b_1\}) + .. + P(\{b_k\}) = k \cdot P(\{b_1\}) = \frac{\#(A)}{\#(S)}$

**Lemma:** Let S be a finite sample space- Pis the uniform probability then for each $AcS$.

$P(A) = \frac{\#(A)}{\#(S)}$.

**Eg:** Let a fair due be rolled twice. A is the event in which a common number was obtained in the 2 trials.

The sample space of this experiment is $S = \{1, 2, 3, 4, 5, 6\}^2$

$A = \{(1, 1), (2, 2), ... , (6, 6)\}$

$P(A) = \frac{\#(A)}{\#(S)} = \frac{6}{36} = \frac{1}{6}$

Let B be the event of obtaining a sum of 5.

$B = \{(1, 4), (2, 3), (4, 1), (3, 2)\}$

Again the probability of B is

$P(B) = \frac{4}{36} = \frac{1}{9}$

**Eg:** (Monty hall) a prize is hidden behind 1 of 3 doors. A player picks one box, then another box is revealed not to be the prize. The player is then given an opportunity to stick or switch.

**Soln:** Let the door first picked be door 1. The others are door 2 and door 3. The sample space is $S = \{1, 2, 3\}$.

Stick: In the case where he wins by staying. Winning event is a singleton = $\{1\}$, so $P(\{1\}) = \frac{1}{3}$.

Switch: In the case where he wins by switching. Winning event is a doubleton $= \{2, 3\}$, so $P(\{2, 3\}) = \frac{2}{3}$

## 17-09-2019

**Def:** Let A,B be events. The probability of A given B is P(A|B)=$\frac{P(A \cap B)}{P(B)}$, if P(B)>0.

**Eg:** A fair die is rolled twice. Compute the probability that a sum of 6 is obtained, while the difference is equal to 4.

**Soln:** Let $S = \{1, 2, ..., 6\}^2$ be the sample space of this experiment.

$A = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$ is the event in which a sum of 6 is obtained.

$B = \{(1, 5), (2, 6), (6, 2), (5, 1)\}$ is the event in which a difference of 4 is obtained.

Then the solution is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{(1, 5), (5, 1)}{P(B)} = \frac{\frac{2}{36}}{\frac{4}{36}} = \frac{1}{2}$$

Remark: The computation of P(A|B) can be regarded as the probabilty of A∩B as an event in a smaller sample space B

**Eg:** A family with 2 children is picked at random. Find the probability that the youngest kid a boy, given that the oldest is a boy.
**Sol**n: Let S={bb,bg,gb,gg} be the sample space.
$A = \{bb, gb\}$. Goal: event where the youngest is a boy.
$B = \{bb, bg\}$. Condition: event where the eldest is a boy.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{bb\})}{P(B)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$$

**Eg:** A family with 2 children is picked at random. Find the probability that the remaining kid is a boy given that the that one child is a boy.
Soln: $S = \{bb, bg, gb, gg\}$
Let $A = \{bb\}$
Let $B = \{bb, bg, gb\}$

$$P(A|B) = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

**Eg:** There are $m$ red balls and $n$ green balls in a bag. Calculate the chance of getting 2 red balls, when 2 balls are drawn without replacement.
**Soln**: Let A be the event in which 2 red balls are drawn, B is the event in whic a red ball was obtained in the 1.st draw.
$P(A) = P(A \cap B) = P(A|B) \cdot P(B)$, $P(B) = \frac{m}{m+n}$ and
$P(A|B) = \frac{m-1}{m+n}$, given that the sample space is now after one has been drawn in advance, so one less red ball to draw.
$$P(A) = \frac{m}{m+n} \cdot \frac{m-1}{m+n}$$

Principle of multiplication:
$P(A \cap B \cap C) = P(C|B \cap A)P(B \cap A|A)P(A)$

**Eg:** A fair die is rolle for twice. Compute the probability of obtaining a sum of 5, given that 3 was obtained n the 1.st trial.
Soln: $S = \{1, 2 ..., 6\}^2$ is the sample space.
$A$ ={ $(1, 4.(2, 3), (3, 2), (4, 1)$ }
$B = \{(3, 1), (3, 2), (3, 3),..., (3, 6)\}$ , 3 first.
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{6}$$
$$P(A) = \frac{4}{36} = \frac{1}{9}$$
So the knowledge of the occurence of B, enchance the occurence of A.

**Eg:** (Counter) A fair die is rolled for twice. Compute the probability of getting a sum of 7, given that 3 was obtained in the 1.st trial.
**Soln:** $S = \{1, 2,..., 6\}$

$A = \{(1,6), (2,5),..., (6,1)\}$. 6 options.
$B = \{(3,1), (3,2) ..., (3,6)\}$. 6 options

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{6}$$
$$P(A) = \frac{6}{36} = \frac{1}{6}$$

The occurence of B neither enchance nor supress the chance of A occuring.

$$\frac{0.8 \cdot 0.2}{0.8}$$

$$0.2000000000 \tag{2.2.1}$$

**Def.** Two events A,B are independent if $P(A \cap B) = P(A) \cdot P(B)$
**Cor:** Let $P(B) > 0$, then A, B are independent events if and only if $P(A) = P(A|B)$

**Eg:** Let a coin flip twice.
A head was obtained 1st. then a tail in the 2nd. Seems independent.
A head was obtained 1st. then a tail again in the 1st. They are dependent.

# 19-09-2019

**Remember;**

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

## Dependence

**Def:** A,B,C are independent if $P(A \cap B \cap C) = P(A)P(B)P(C)$
**Remark:** In general. The **independence of several events** is **different** from their **pairwise independence.**

**Def:** Let S be the sample space of en experiment. The events $S_1, S_2 ...$ form a partition if:
- $S_i \cap S_j = \varnothing$ for all $i \neq j$.
- $S_1 \cup S_2 \cup ... = S$

**Thm:** (Bayes) If $S_1, S_2, ...$ is a set of events forming a partition of the sample space of an experiment.
Then for every event A: $P(A) = \sum_j P(A|S_j) \cdot P(S_j)$

**Proof:** $P(A) = P(A \cap S)$
$= P(A \cap (S_1 \cup S_2 \cup ...))$
$= P((A \cap S_1) \cup (A \cap S_2) \cup ...)$

Verify that if $i \neq j$, then $(A \cap S_i) \cap (A \cap S_j) = (A \cap A) \cap S_i \cap S_j = A \cap (S_i \cap S_j) = \varnothing$
$P((A \cap S_1) \cup (A \cap S_2) \cup ...)$

$=P(A\cap S_1)+P(A\cap S_2)+\ldots$
$=P(A|S_1)P(S_1)+P(A|S_2)P(S_2)+\ldots$

**Eg:** A fair die is rolled, then a fair coin is flipped n times, where n is the value of the die roll.
**Goal:** Calculate the probability of obtaining 4 heads.
**Soln:** For each $i$, let $S_i$ be the event in which $i$ was obtained when rolling the die.

Partition:
- $S_i \cap S_j = \varnothing$ for all $i \neq j$ : Rolling 1 die does not give a chance for multiple rolls, so no overlap. They are piecewise disjoint.
- $S_1 \cup S_2 \cup \ldots = S$.: Yes, $S_1 \cup S_2 \cup .. S_6 = S$

$P(A)=P(A|S_1)P(S_1)+P(A|S_2)P(S_2)+\ldots P(A|S_6)P(S_6)$, However to get 4 heads, only $i \geq 4$ is usable.

$=P(A|S_4)P(S_4)+P(A|S_5)P(S_5)+P(A|S_6)P(S_6)$

$$\frac{1}{16}\cdot\frac{1}{6}+\frac{5}{32}\,\frac{1}{6}+\frac{15}{64}\,\frac{1.}{6}$$

$$0.07552083333 \tag{3.1.1}$$

**Eg:** A certain defect is present in 1% of the population. A test is used, with the probability of correctly obtaining a positive result is 90%. So 10% for a false positive.
Goal: Compute the probability that a person has such defect, given a positive result.
**Soln:** Let A be the event that random person has the defect
and let B be the event that this random person gets a positive result.

$P(B)=P(B|A)\cdot P(A)+P(B|A^c)+P(A^c)$ by use of Bayers theorem. Where $P(B|A^c)$ is the probability of a false positive.

$=(0.9)\cdot(0.01)+(0.1)\cdot(0.99)$

So $P(A|B)=\dfrac{P(A\cap B)}{P(B)}=P(B|A)\,\dfrac{p(A)}{P(B)}$

$$=\frac{(0.9)\cdot(0.01)}{(0.9)\cdot(0.01)+(0.1)\cdot(0.99)}=\frac{1}{12}$$

**Def:** Let S be the sample space, where a random variable is a function from S to $\mathbb{R}$.

**Eg:** Let a die be rolled twice.So that its sample space is $S=\{1,2,\ldots,6\}^2$.
Define: $X: S\to\mathbb{R}$. A function $X$ which brings a value from the sample space $S$ into the real space $\mathbb{R}$.
*Example*: $X((x,y))=x$ for all $(x,y)\in S$. It is a random variable.
$x$ is also known as the number, obtained in the 1st. trail.

Also: $Y: S\to\mathbb{R}$ defined by $Y((x,y))=y$ for all $(x,y)\in S$.
$y$ is a random variable known as the number obtained in the 2nd. trial.

Additionally: a function is given $T: S\to\mathbb{R}$ is defined by $T((x,y))=x+y$ for all $(x,y)\in S$.
is a random variable known as "the sum of obtained values".

Additionally: a function is given $M: S\to\mathbb{R}$ is defined by $M((x,y))=\max\{x,y\}$ for all $(x,y)\in S$.
is a random variable known as the "largest number(max.) obtained"

**Eg:** Let a coin be flipped 5 times, so that its sample space is $\{H, T\}^5$.
Define $X : S \to \mathbb{R}$ by
$X((a, b, c, d, e)) = \textit{the number of members } \textbf{in } (a, b, c, d, e) \textit{ which are equal } \textbf{to } H$

## 24-09-2019 Probabilities

**Def:** A random variable is a function from a sample to $\mathbb{R}$.

**Def:** Let $X, Y : S \to \mathbb{R}$ be random variables.
1) $X + Y : S \to \mathbb{R}$ is defined by
$(X + Y)(s) = X(s) + Y(s)$ for all $s \in S$.

2) $XY : S \to \mathbb{R}$ is defined by $(XY)(s) = X(s)Y(s)$ for all $s \in S$.

Where $(s)$ is the variable of function $X$ and $Y$.

**Eg:** Let a die be rolled for twice.
X is the number obtained in the 1st. trial.
Y is the number obtained in the 2nd. trial.
T is the sum obtained. Then T=X+Y.

**Def:** Let $X : S \to \mathbb{R}$ be a random variable and $a, b \in \mathbb{R}$.
- $P(X = a) = P(\{s \in S : X(s) = a\})$
- $P(a \leq X \leq b) = P(\{s \in S : a \leq X(s) \leq b\})$
- $P(X < a) = P(\{s \in S : X(s) < a\})$

**Eg:** Let a fair die be rolled. X is the number obtained.
Calculate $P(X \leq 3.14)$.
**Soln:** Let $S = \{1, 2, ..., 6\}$ be the sample space, so:
$P(X \leq 3.14) = P(\{s \in S : X(s) \leq 3.14\})$
Consider the fact that an element of the sample space gives a value of that element in this case (Might result in other values given on the event condition).
So {1,2,3} is favorable to the event.
$P(\{1, 2, 3\}) = \frac{1}{2}$

**Def:** Let X be a random variable. The cumulative distributive function of X is a function $F : \mathbb{R} \to \mathbb{R}$ by
$F(t) = P(X \leq t)$ for all $t \in \mathbb{R}$.

**Lemma:** The commulative distribution function of a random variable is increasing.
**Proof:** Let $X : S \to \mathbb{R}$ be a random variable. $F : \mathbb{R} \to \mathbb{R}$ is its commulative distribution. Then for each $t_1 < t_2$

$\{s \in S : X(s) \le t_1\}$ $\{s \in S : X(s) \le t_2\}$
$F(t_1) = P(\{s \in S : X(s) \le t_1\}) \le P(\{s \in S : X(s) \le t_2\}) = F(t_2)$

**Def:** A random variable X is dicrete if there exist $a_1, a_2, \ldots \in \mathbb{R}$ such that $\sum_j P(X = a_j) = 1$

**Def:** If X is a discrete random variable, its probability mass function (or just its probability distribution function) is the function $f : \mathbb{R} \to \mathbb{R}$ defined by
$f(t) = P(X = t)$ for all $t \in \mathbb{R}$.

**Eg:** Let X be the number obtained when a fair die is rolled.

**Soln:** So that its mass function is $f(t) = \frac{1}{6}$ if $t \in \{1, 2, \ldots, 6\}$ or $f(t) = 0$ otherwise.

$P(X = 1) = \frac{1}{6}$
$P(X = 1.5) = 0$
$P(X = 2) = \frac{1}{6}$
...
$P(X = 6) = \frac{1}{6}$
Let its cummulative distribution be F.
$F(0.2) = 0$,
$F(1) = \frac{1}{6}$, $F(1.4) = \frac{1}{6}$
$F(2) = \frac{2}{6}$

**Eg:** Let X be the sum obtained whn a fair die is rolled for twice. Compute its mass function $f$.
**Soln:**
$P(X = 2) = \frac{1}{36}$
$P(X = 3) = \frac{2}{36}$
...
$P(X = 6) = \frac{5}{36}$
$P(X = 7) = \frac{6}{36}$ ----
$P(X = 8) = \frac{5}{36}$
...
$P(X = 12) = \frac{1}{6}$

$$f(t) = \left\{ \left[ \begin{array}{ll} \frac{(6 - |t - 7|)}{36}, & \textbf{if } t \in \{2, 3, \ldots, 12\} \\ 0 & , otherwise \end{array} \right] \right.$$

**Def:** Let X be a random variable . Its probability density function is a function $f: \mathbb{R} \to \mathbb{R}$ so that for each $a \le b$,

$$P(a \le X \le b) = \int_a^b f(t)\, \mathrm{d}t$$

**Def:** A random variable is continuous if its density exists.

**Lemma:** If X is a continuous random variable. $P(X=a) = 0$ for all a, since the integral for integrands equal to eachother is 0.

**So**, the following methods are good to describe the following each:
Discrete random variables: Cumulative distribution and Mass function
Continuous random variable: Cumulative distribution and density.

**Lemma:** if f is the density of a random variable, then

$$\int_{-\infty}^{\infty} f(t)\, \mathrm{d}t = 1$$

**Lemma:** if $f$ is the density of a random variable, then
$f(t) \ge 0$ for almost all t.

**Thm:** if $F$, $f$ are the cumulative distribution and density of a random variable respectively. Then $F' = f$

# 26-09-2019

**Def:** Let $n \in \mathbb{N}$, $0 < p < 1$, A discrete random variable X follows the binomial distribution with size n and probability p if its mass function is

$$p(X=k) = \begin{cases} \left[\left[\begin{array}{c} n \\ k \end{array}\right] \cdot p^k \cdot \left(1-p\right)^{n-k}, \textbf{ if } k \in \left\{0, 1, 2 \ldots, n\right\} \\ 0 \text{ , } otherwise \end{cases}\right], \text{ where } \left[\begin{array}{c} n \\ k \end{array}\right] = \frac{n!}{k!(n-k)!}$$

Remark the binomial thm: For every $n \in \mathbb{N}$, $x, y \in \mathbb{R}$, then

$$(z+y)^n = \sum_{k=0}^{n} \left[\begin{array}{c} n \\ k \end{array}\right] x^k \cdot y^{n-k}$$

Applied in case:
$x = p$, $\mathrm{y} = 1 - p$

$$1 = (p + (1-p))^n = \sum_{k=0}^{n} \left[\begin{array}{c} n \\ k \end{array}\right] p^k \cdot (1-p)^{n-k} = \sum_{k=0}^{n} P(X=k)$$

Remark: A Bernoullis' trial is an experiment with 2 possible outcomes:
Succes or failure, with probabilities p, 1-p respectively.
Repeat the trial above for n times, let X be the number of successes in n trials.
$P(X=0) = (1-p)^n$, 0 succes
$P(X=1) = n \cdot p \cdot (1-p)^{n-1}$, 1 succes

$P(X=2) = \begin{bmatrix} n \\ 2 \end{bmatrix} \cdot p^2 (1-p)^{n-2}$, Thm: Let S be a set with n members. then there are $\begin{bmatrix} n \\ k \end{bmatrix}$ subsets of S consisting of $k$ members.

In general $P(X=k) \begin{bmatrix} n \\ k \end{bmatrix} p^k (1-p)^{n-k}$, so that X is binomial distributed.

$(0.65\cdot 0.4 + 0.5\cdot 0.6) - (0.65\cdot 0.4 + 0.5\cdot 0.6)\cdot 0.15 + 0.15\cdot 0.8$

$$0.59600 \tag{5.1}$$

Def: Let $\mu \in \mathbb{R}$ and $\sigma > 0$. A random variable X follows the normal distribution with mean $\mu$ and standard deviation $\sigma$ if its density is $f(x) = \frac{1}{\sqrt{2\cdot\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}$ for all $x \in \mathbb{R}$

Remark: it can be shown that $\frac{1}{\sqrt{2\cdot\pi}\,\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}} \, dx = 1$

Lemma: If X us normally distributed, so is $a\cdot X + b$ for all $a, b \in \mathbb{R}$.

Lemma: If X is normally distributed with mean $\mu$ and standard deviation $\sigma$, then $\frac{(X-\mu)}{\sigma}$ is standard normally distributed.

Proof: Let $Y = \frac{(X-\mu)}{\sigma}$, For each $y \in \mathbb{R}$.

$P(Y \le y) = P\left(\frac{(X-\mu)}{\sigma} \le y\right)$

$= P(X \le \mu + \sigma y)$

$= \frac{1}{\sqrt{2\cdot\pi}\,\sigma} \cdot \int_{-\infty}^{\mu+\sigma y} e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}} \, dx$, use substitution by letting $t = \frac{x-\mu}{\sigma}$, $dt = \frac{1}{\sigma} dx$, so when $x \to \infty$ then $t \to \infty$ and when $x = \mu + \sigma y$ then $t = y$.

$= \frac{1}{\sqrt{2\cdot\pi}} \cdot \int_{-\infty}^{y} e^{-\frac{t^2}{2}} \, dx$, which is the commulative distribution of Y.

Its density is $\frac{1}{\sqrt{2\cdot\pi}} \cdot e^{-\frac{t^2}{2}} = f(t)$, which is the density of the standard normal.

Def: If X is a discrete random variable with mass function $f$. Its expectation is $E(x) = \sum_k P(X=k)\cdot k$

If X is continuous with density $f$. Its expectation is $E(X) = \int_{-\infty}^{\infty} x\cdot f(x) \, dx$

Notice: The average and expectation value is different.

# Review of Chapter 2 - 27-09-2019

## Morgan's Law and Law of total probability and Beyes Theorem.

**Morgan's Law:**

$(A \cap B)^c = A^c \cup B^c$, $\left(A^c \cup C\right)^c = A^c \cap B^c$

**Law of total probability:**

$(A \cap B)$ and $\left(A \cap B^c\right)$ are disjoint.

$A = (A \cap B) \cup \left(A \cap B^c\right)$

$P(A) = P\left(A \middle| B\right) \cdot P(B) + P\left(A \middle| B^c\right) \cdot P\left(B^c\right) \rightarrow B_1, B_2, \ldots, B_k$ are exclusive.

$$P(A) = \sum_{i=1}^{k} P\left(A \middle| B_i\right) \cdot P\left(B_i\right)$$

**Beye's Theorem:**

$$P\left(B \middle| A\right) = \frac{P(B) \cdot P(A|B)}{P(A)}$$

And Law of probability and Beye's theorem can be combined:

$$P\left(B \middle| A\right) = \frac{P(B) \cdot P(A|B)}{P\left(A \middle| B\right) \cdot P(B) + P\left(A \middle| B^c\right) \cdot P\left(B^c\right)}$$

# Forventningsværdi 03-10-2019

Def:

Let X be a discrete random variable with mass function F

$$E(X) = \sum_{X} X \cdot F(X)$$

If $X$ is a continuous random variable with density $f$

$$E(X) = \int_{-\infty}^{\infty} X \cdot f(X)\, \mathrm{d}X$$

Eg: Let X be the number obtained when a fair die is rolled. Compute E(X)

Soln: The mass fnc. of X is $P(X=1) = P(X=2) = \ldots = P(X=6) = \frac{1}{6}$

$$E(X) = \frac{1\cdot 1}{6} + \frac{2\cdot 1}{6} + \ldots + \frac{6\cdot 1}{6} = \frac{7}{2}$$

Remark: in general if X is uniformly distributed over {1,2,...,n}, then $E(X) = \frac{1}{2}(n+1)$

Eg: Let M be the maximum number obtained when a fair die is rolled for twice. Compute E(M).

Soln: $P(M=1) = \left(\frac{1}{6}\right)^2$

$$P(M=2) = \left(\frac{3}{6}\right)^2 = \frac{(2^2-1^2)}{6^2},\ P(M=3) = \frac{(3^2-2^2)}{6^2},\ldots,\ P(M=6) = \frac{(6^2-5^2)}{6^2}$$

$$E(M) = 1\cdot\frac{1}{6^2} + \frac{2\cdot(2^2-1^2)}{6^2} + \ldots + \frac{6\cdot(6^2-5^2)}{6^2}$$

$$E(M) = \frac{1}{6^2}(-2^2-3^2-4^2-5^2) + \frac{1}{6^2} + 6.$$

$$E(M) = 4.527777778 \quad \textbf{(7.1)}$$

**Continuous:**

**Eg:**

Let X be uniformly distributed on [0,1]. Compute E(X).

**Soln:** The density of X is $f(x)\left\{\left[\begin{array}{ll} 1, & \textbf{if } 0 \le X \le 1 \\ 0, & \textit{otherwise} \end{array}\right]\right.$

$$E(X) = \int_{-\infty}^{\infty} X\cdot f(X)\,\mathrm{d}X = \int_0^1 X\cdot 1\,\mathrm{d}X = \left[\frac{1}{2}X^2\right]_0^1 = \frac{1}{2}$$ as expected.

**Eg:**

Let X be the uniformly distributed on [0,1]. Let $Y=X^2$. Compute E(Y)

**Soln**: For each $0 \le t \le 1$ , $P(Y \le t) = P(X^2 \le t) = P(X \le \sqrt{t}) = \sqrt{t}$

So if $0 < t \le 1$, $\frac{d}{dt}P(Y \le t) = \frac{1}{2\cdot\sqrt{t}}$

$$E(Y) = \int_0^1 t\cdot\frac{1}{2\cdot\sqrt{t}}\,\mathrm{d}t = \frac{1}{2}\int_0^1 \sqrt{2}\,\mathrm{d}t = \left[\frac{1}{3}\cdot t^{\frac{3}{2}}\right]_0^1 = \frac{1}{3}$$

Let $X: S \to \mathbb{R}$ be a random variable, $g: \mathbb{R} \to \mathbb{R}$ is a function.
Then $g(X): S \to \mathbb{R}$ is defined by
$g(X)(s) = g(X(s))$ **for** *all* $s \in S$.

**Thm:** Let X be a continuous random variable with density $f$.

$g: \mathbb{R} \to \mathbb{R}$ is a function. Then $E(g(X)) = \int_{-\infty}^{\infty} g(X)\cdot f(X)\,\mathrm{d}X$

**Pf:** For each t, $P(g(X) \le t) = P(X \le g^{-1}(t)) = \int_{-\infty}^{g^{-1}(t)} f(x)\,\mathrm{d}x$

Assume that g is increasing. If not divide the space into discrete amounts wherein g is either decreasing or increasing.

Density:

$$\frac{d}{dt}P(g(X) \le t) = f\left(g^{-1}(t)\right)\cdot\left(g^{-1}(t)\right)'$$

Therefore $E(g(X)) = \int_{-\infty}^{\infty} t\cdot f\left(g^{-1}(t)\right)\cdot\left(g^{-1}(t)\right)' \mathrm{d}t$

Use substitution, so $u = g^{-1}(t)$ so that $du\left(\left(g^{-1}(t)\right)'dt\right.$

$E(g(X)) = \int_{-\infty}^{\infty} g(u)\cdot f(u)\cdot du$

**Thm:** If $X$ is a discrete random variable with mass function f, $g:\mathbb{R}\to\mathbb{R}$ is a function. Then $E(g(X)) = \sum_x g(x)\cdot f(x)$

**Remark:** In general if X is uniformly distributed over {1,2,...,n},
$E\left(X^2\right) = \frac{1}{6}(n+1)(2n+1)$

**Eg:** Compute $E\left(X^2\right)$ if X is uniformly distributed over [0,1].

$E\left(X^2\right)\int_0^1 X^2\cdot 1\,\mathrm{d}x = \frac{1}{3}$

**Corolar:** Let X be a random variable, a $\in\mathbb{R}$, $E(a\cdot X) = a\cdot E(X)$

**Def:** Let X be a random variable, Var(X) is $E\left((X - E(X))^2\right)$

**Lemma:** Var(X)≥0 for all X.

**Def:** $\sigma = \sqrt{Var(X)}$ is called the standard deviation of X.

**Lemma:** $Var(X) = E\left(X^2\right) - E(X)^2$

***Proof*:** $Var(X) = E\left(X^2\right) - E(X)^2$

$Var(X) = E\left((X - E(X))^2\right)$

$= E\left(X^2 - 2XE(X) + E(X)^2\right.$

$= E\left(X^2\right) - E(2XE(X)) + E\left(E(X)^2\right)$

$= E\left(X^2\right) - 2E(X)E(X) + E\left(X^2\right) = E\left(X^2\right) - E(X)^2$

**Eg:** Compute Var(X) if X is uniformly distributed over $\{1, 2, 3 ..., n\}$

**Soln:**

$Var(X) = E\left(X^2 - E(X)^2\right) = \frac{1}{6}(n+1)(2n+1) - \left(\frac{1}{2}(n+1)\right)^2 = \frac{1}{12}(n+1)(4n+2-3n - 3) = \frac{1}{12}(n+1)(n-1)$

**Eg:** Compute $Var(X)$ if X is uniformly distributed over [0,1].

$Var(X) = E\left(X^2\right) - E(X)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

**Lemma:** If X is a random variable, $a \in \mathbb{R}$ $Var(X+a) = Var(X)$
**Pf:** $E(X+a) = E(X) + E(a) = E(X) + a$
$Var(X+a) = E\left((X+a-E(X+a))^2\right) = E\left((X+a-E(X)-a)^2\right) = E\left((X-E(x))^2\right) = Var(X)$

**Lemma:** If X is a random variable , $a \in \mathbb{R}$, $Var(aX) = a^2 Var(X)$
**Pf:** $E(aX) = aE(X)$ .
$Var(aX) = E\left((aX-E(aX))^2\right) = E\left((aX-aE(x))^2\right) = E\left(a^2(x-E(x))^2\right) = a^2 \cdot E\left(a^2(x-E(x))^2\right)$
$= a^2 \cdot Var(X)$

**Eg:** Calculate E(X) if X is uniformly distributed over $[a, b]$
Soln: $\frac{x-a}{b-a}$ is uniform over $[0, 1]$, $\frac{1}{(b-a)^2} Var(X) = Var\left(\frac{x-a}{b-a}\right) = \frac{1}{12}$

# 08-10-2019

**Pf:** If X follows the binormial distribution with parameters $n, p$.

Then $P(X=k) = \left[\begin{matrix} n \\ k \end{matrix}\right] p^k (1-p)^{n-k}$ for $k \in \{0,..., n\}$

Then

$$E(X) = \sum_{k=0}^{n} k \left[\begin{matrix} n \\ k \end{matrix}\right] p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} \cdot n \cdot p^k (1-p)^{n-k}$$

$= \sum_{k=1}^{n} \left[\begin{matrix} n-1 \\ k-1 \end{matrix}\right] \cdot n \cdot p^k (1-p)^{n-k}$ , Recall: $(X+Y)^m = \sum_{k=0}^{m} \left[\begin{matrix} m \\ k \end{matrix}\right] \cdot x^k \cdot y^{m-k}$, so

$$= \sum_{k=0}^{n-1} \left[\begin{matrix} n-1 \\ k \end{matrix}\right] \cdot n \cdot p^{k+1} (1-p)^{n-k-1}$$

$$= n \cdot p \sum_{k=0}^{n-1} \left[\begin{matrix} n-1 \\ k \end{matrix}\right] \cdot p^k (1-p)^{n-1-k} = n \cdot p \cdot (p+1-p)^{n-1}$$

$$= n \cdot p$$

**Thm:** If X follows the binomial distribution with parameters $n, p$. Then $E(X) = n \cdot p$

**Thm:** If X is normally distributed with mean $\mu$ and standard deviation $\sigma$, then $E(X) = \mu$ and $Var(X) = \sigma^2$

**Remember:** $\int_a^b u \cdot v' = [u \cdot v]_a^b - \int_a^b u' \cdot v$

If X is normally distributed, then $\frac{(X-\mu)}{\sigma}$ is standard normally distributed:

$E(X) = E\left(\sigma \cdot \frac{X-\mu}{\sigma} + \mu\right) = \sigma \cdot E\left(\frac{X-\mu}{\sigma}\right) + \mu = \mu$

$Var(X) = Var\left(\sigma \cdot \frac{X-\mu}{\sigma} + \mu\right)$, if one has a variable + a constant, then disregard the constant, so

$Var(X) = Var\left(\sigma \cdot \frac{X-\mu}{\sigma}\right) = \sigma^2 \cdot Var\left(\frac{(X-\mu)}{\sigma}\right) = \sigma^2$

**Thm:** If X,Y are random variables, $E(X+Y) = E(X) + E(Y)$

**Eg:** Let $Xj$ be the number of success obtained when a Bernowlli's trial is alone for the $j$th time. Then $X = X_1 + X_2 + \ldots + X_n$ is binormial distributed with parameters $n, p$.

$E(X) = E(X_1 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$

$= 1 - p + 0 \cdot (1-p) + \ldots = n \cdot p$

**Eg:** Let N be the minimum number obtained when a fair die is rolled twice. Calculate $E(N)$.
**soln:** Let X,Y be the number obtained in the 1st. 2nd trial respectively. M is the max. number obtained. Then

$M + N = X + Y \rightarrow E(M+N) = E(X+Y)$

$E(N) = E(X) + E(Y) - E(M)$

**Def:** Two random variable $X, Y: S \rightarrow \mathbb{R}$ are independent. if for every $a, b, c, d \in \mathbb{R}$.
$\{s \in S : a \leq X(s) \leq b\}$ and $\{s \in S : c \leq Y \leq d\}$ are independent.

**Thm:** If $X$, $Y$ are independent random variables, then $E(X \cdot Y) = E(X) \cdot E(Y)$

**Cor:** If $X$, $Y$ are independent variables, then $Var(X+Y) = Var(X) + Var(Y)$

**Eg:** Let $X_j$ be the # of succes in the $j$th Bernoulli's trial. $X(X_1 + X_2 + \ldots + X_n)$ is binormial distributed with parameters $n, p$.

$Var(X) = Var(X_1 + X_2 + \ldots + X_n)$

$= Var(X_1) + Var(X_2) + \ldots + Var(X_n)$, since $X_1 \ldots X_n$ are independent. Notice all variance are equal.

$= n \cdot p(1-p)$

$Var(X_1) = E(X_1^2) - E(X_1)^2 = E(X_1) - E(X_1)^2$, since $X_1$ is a boolean, so $X$ to the power of anything is just $X$.

$Var(X_1) = p \cdot (1-p)$

# 10-10-2019

**Def:** Let X be a random variable, its moment generating function is a function $f: \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(t) = E(e^{t \cdot x})$ for all $t$.

**Cor**: If $f$ is the moment generating function of a random variable, then $f(0) = 1$

**Eg:** Let X follow the binormial distribution with parameters $n, p$.

Let $X_j$ be the number of succes' obtained when a Bernoulli's trial with prob of succes $p$ is done at time $j$.
Then $X = X_1, X_2 + ... + X_n$. The moment generating function $f$ for X is

$$f(t) = E\left(e^{t \cdot X}\right) = E\left(e^{t \cdot \left(X_1, X_2 + ... + X_n\right)}\right) = E\left(e^{t \cdot X_1} \cdot e^{t \cdot X_2} \cdot ... \cdot e^{t \cdot X_n}\right)$$

$$= E\left(e^{t \cdot X_1}\right) \cdot E\left(e^{t \cdot X_2}\right) \cdot ... \cdot E\left(e^{t \cdot X_n}\right)$$ , as $X_1, X_2, ..., X_n$ are independent random variables.

**Identically distributed:** Has the same mass function. In this case it's $X_1$ with $X_2$ and so on, since they are booleans

$$= E\left(e^{t \cdot X_1}\right)^n$$

$$= \left(p \cdot e^t + (1 - p) \cdot 1\right)^n = \left(p \cdot e^t + 1 - p\right)^n$$

**Eg:** Let Z be standard normally distributed. Its moment generating function is $E\left(e^{t \cdot Z}\right)$
Density of standard normal distrubution =

$$\frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{-\infty} e^{t \cdot x} \cdot e^{-\frac{x^2}{2}} \, dx = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{-\infty} \cdot e^{-\frac{x^2}{2} + t \cdot x - \frac{t^2}{2} + \frac{t^2}{2}} \, dx$$

$$= \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\frac{t^2}{2}} \int_{-\infty}^{-\infty} \cdot e^{-\frac{(x-t)^2}{2}} \, dx$$

$$= e^{\frac{t^2}{2}}$$ , as $\dfrac{e^{-\frac{(x-t)^2}{2}}}{\sqrt{2 \cdot \pi}}$ is the density of the normal distribution with mean t and variance 1.

So, for a continous $x$ with density $\rho$. $f(t) = E\left(e^{t \cdot x}\right) = \int_{-\infty}^{\infty} e^{t \cdot x} \cdot \rho(x) \, dx$

$$\frac{d}{dt} f(t) = \int_{-\infty}^{\infty} \frac{d}{dt}\left(e^{t \cdot x} \cdot \rho(x)\right) dx = \int_{-\infty}^{\infty} e^{t \cdot x} \cdot x \cdot \rho(x) \, dx = E\left(X \cdot e^{t \cdot x}\right)$$

**Eg:** Let $X$ be normal distributed with mean $\mu$ and variance $\sigma^2$.
Its moment generating function is

$$E\left(e^{t \cdot x}\right) = E\left(e^{t \cdot (X - \mu)} \cdot e^{t \cdot \mu}\right) = e^{t \cdot \mu} \cdot E\left(e^{\frac{\sigma \cdot t \cdot (x - \mu)}{\sigma}}\right)$$

$$= e^{\mu \cdot t} \cdot e^{\sigma^2 \cdot \frac{t^2}{2}}$$ , given by the last paragraph.

**Thm:** Let $X$ be a random variable. $f$ is its moment generating function. Then for each integer $k$.
$E\left(X^k\right) = f^{(k)}(0)$, $E\left(X^k\right)$ are the "moments of $X$".

**Eg:** Let $X$ follow the binormial distribution with parameters $n, p$.
Its moment generating function is $f(t) = \left(p \cdot e^t - p + 1\right)^n$

$$f'(t) = n \cdot \left(p \cdot e^t - p + 1\right)^{n-1} \cdot p \cdot e^t$$

$$f''(t) = n \cdot (n - 1) \cdot \left(p \cdot e^t - p + 1\right)^{n-2} \cdot p^2 \cdot e^{2 \cdot t} + n\left(p \cdot e^t - p + 1\right)^{n-1} \cdot p \cdot e^t$$

$$E(X) = f'(0) = n \cdot p$$

$E(X^2) = n\cdot(n-1)\cdot p^2 + n\cdot p$ , so that
$Var(X) = E(X^2) - E(X)^2 = n\cdot(n-1)\cdot p^2 + n\cdot p - n^2\cdot p^2$

**Eg:** Let Z be standard normally distributed. It's moment generating function is $f(t) = e^{\frac{t^2}{2}}$.

$f'(t) = t\cdot e^{\frac{t^2}{2}}$

$f''(t) = (t^2+1)\cdot e^{\frac{t^2}{2}}$

$f'''(t) = (t^3+3t)\cdot e^{\frac{t^2}{2}}$

$f^{(4)}(t) = (t^4+3\cdot t^2+3t^2+3)\cdot e^{\frac{t^2}{2}} = (t^4+6\cdot t^2+3)\cdot e^{\frac{t^2}{2}}$

$E(Z) = 0$, plug 0 into $f'(t)$
$E(Z^2) = 1$, plug 0 into $f''(t)$ and so on.
$E(Z^3) = 0$
$E(Z^4) = 3$

**?:** Let $X$ be a random variable with $E(X) = \mu$, $Var(X) = \sigma^2$
$E\left(\left(\frac{(X-\mu)}{\sigma}\right)^3\right)$.
What defines the sign of this function? Is the expectation that of a negative or positive number?
**Ans:** When X is likely to be a lot greater than μ, rather than a lot less than μ.

**Def:** If $E(X) = \mu$, $Var(X) = \sigma^2$, then the **skewness** of $X$ is $E\left(\left(\frac{(X-\mu)}{\sigma}\right)^3\right)$

$X$ is skewed to the right if $E\left(\left(\frac{(X-\mu)}{\sigma}\right)^3\right) > 0$. Reverse inequality for left.

**?:** For the 4. moment, what makes $E\left(\left(\frac{(X-\mu)}{\sigma}\right)^4\right) > 3$?

**Ans:** When the density of $X$ is not locally concentrated at the center of standard distribution. So, the function is more level on a global level.

**Def:** If $E(X) = \mu$, $Bar(X) = \sigma^2$, the Kurtosis of $X$is $E\left(\left(\frac{(X-\mu)}{\sigma}\right)^4\right)$. The excessive Kurtossis of $X$ is $E\left(\left(\frac{(X-\mu)}{\sigma}\right)^4\right) - 3$.

**Remark:** Its purpose is to compare its tail part with that of the standard normal. So, it compares the distribution of density or how level the density is.

**Thm: Law of large number:** Let $X$ be a random variable with $E(X) = \mu$.

Then $P\left(\left|\frac{(X_1+X_2+...+X_n)}{n}-\mu\right|=0\right) \approx 1$, so the probability of the average $x_i$ minus $\mu$ absolute being equal to 0 is very high.
Where $(X_1+X_2+...+X_n)$ are independent random variables, which are distributed in the same way as $x$.

**Thm: Central Limit Thm:** Let $X_1, X_2, ..., X_n$ be "iids"(Independent and identically distriuted random variables) and $n$ is large.

Then $\frac{\frac{(X_1+X_2+...+X_n)}{n}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is approx. standard normally distributed.

# 15-10-2019

**Thm: Central Limit Thm:** Let $X_1, X_2, ..., X_n$ be "iids"(Independent and identically distriuted random variables) and $n$ is large.

Then $\frac{\frac{(X_1+X_2+...+X_n)}{n}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is approx. standard normally distributed.

Note that: $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$

If $E(X_1)=\mu$, $Var(X1)=\sigma^2$. Then $E\left(\frac{X_1-\mu}{\sigma}\right)=0$, $Var($

**Eg:** There are 100 MC problems in an exam so that there are 5 choices for each problem. Calculate the probability of scoring less than 16 by pure guessing.
**Soln:** Let X be the total score, with a binormial distribution, with parameters Bin(100,1/5)

$$P(X<16)=P(X=0)+P(X=1)+...+P(X=15)=\begin{bmatrix}100\\0\end{bmatrix}\cdot\left(\frac{1}{5}\right)^0\cdot\left(\frac{4}{5}\right)^{100}+\begin{bmatrix}100\\1\end{bmatrix}\cdot\left(\frac{1}{5}\right)^1\cdot\left(\frac{4}{5}\right)^{99}+...+\begin{bmatrix}100\\15\end{bmatrix}\cdot\left(\frac{1}{5}\right)^{15}\cdot\left(\frac{4}{5}\right)^{85}$$

$$evalf\left(\sum_{k=0}^{15}\left(\frac{(100)!}{(k)!(100-k)!}\right)\cdot\left(\frac{1}{5}\right)^k\cdot\left(\frac{4}{5}\right)^{100-k}\right)$$

$$0.1285055148 \quad \textbf{(10.1)}$$

Let $X_j$ be the score obtained in the $j$.th problem.

Then $X_1, ..., X_n$ are iids, and $E(X_1)=\frac{1}{5}$ $Var(X_1)=\frac{4}{25}$

By the Central Limit Theorem,

$$\frac{\frac{X_1+...+X_{100}}{100}-\frac{1}{5}}{\frac{2}{t\cdot\sqrt{100}}}$$ is approx. Standard normally distributed.

$$\frac{\frac{X_1+...+X_{100}}{100}-\frac{1}{5}}{\frac{2}{t\cdot\sqrt{100}}}=\frac{X-20}{4}$$

$$P(X\le 15.5)=P\left(\frac{X-20}{4}\le\frac{15.5-20}{4}\right)=P(StandardNormal\le -1.125)$$

## Statistics

**Def:** A random sample is a collection of iids (Independent identically distributed values).
**Def:** A statistic is a function of a collection of random samples.

**Eg:** Let $f:\mathbb{R}^n\rightarrow\mathbb{R}$ defined by $f(t_1,t_2,...,t_n)=\frac{t_1+t_2+...+t_n}{n}$ for all $t\in\mathbb{R}$ and let $X_1,...,X_n$ be a random sample.

Then $f(X_1,X_2,...X_n)=\frac{X_1+X_2+...+X_n}{n}$ is a statistic, called the average from $X_1$ to $X_n$.

Let $X_1,...,X_n$ be a random sample
Expectation: $E(X)$ is a real number
Average: Is a special kind of random variable called a statistic. If you look at some samples of random numbers then the average is also random and will be different from other experiments. Expectation is always a fixed real number for the data.

**Eg:** Let $X_1$ be the weight of a randomlu picked person in HK.

Then $E(X_1)=\frac{1}{7\cdot 10^6}\cdot weight\ of\ that\ person+\frac{1}{7\cdot 10^6}\cdot weight\ of\ other\ person+...$
$E(X_1)$ = sum of the weights of everybody divided by the total population.
So $E(X_1)$ is also known as the population mean of $X_1$.

Now pick a $n$, so $\frac{X_1+X_2+...+X_n}{n}$, then this is a random variable of $n$ data or a sample of size $n$.
So this is also known as the sample mean or the average.

**Eg:** Let $X_1,...,X_n$ be a random sample

$$s^2=\sum_{j=1}^{n}\frac{\left(X_j-\frac{X_1+...+X_n}{n}\right)^2}{n-1}=\frac{\sum_{j=1}^{n}(X_j-\overline{X})^2}{n-1}$$ is an Eg. of statistic, called the sample

variance.

**Eg:** Let $X_1,..., X_n$ be a random sample
$\max(X_1,..., X_n)$ is a statistic, called the maximum statistic. Same for minimum.
For each $K$, "the $K\,th$ smallest one among $X_1,..., X_n$" is also a statistic.

**Remark:** The objective of the subject statistics is to understand the distributions of statistics(A special kind of random variable).

# 17-10-2019

Thm: Let X,Y follow normal distribution with means $\mu_1, \mu_2$ and variances $\sigma_1^2, \sigma_2^2$ resp. and if X and Y are independent. Then $X + Y$ is normally distributed with mean
$\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$

**Cor:** If $X_1,..., X_n$ is a random sample which is normally distributed with mean $\mu$ and variance $\sigma^2$.
Then $\frac{X_1 + X_2 + ... + X_n}{n}$ is normally distributed with mean $\mu$ and variance

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

Variance is the expected value difference from a result to the actual expected value $E(X)$

**Lemma:** If $X_1,..., X_n$ is a random sample , and n is large (Notice norm. dist. is not assumed). Then
$\frac{X_1 + X_2 + ... + X_n}{n}$ is approx. normally distributed with mean $\mu$ and variance $\frac{Var(X_1)}{n}$
**Pf:** By the central limit thm.

$$\frac{\frac{X_1 + X_2 + ... + X_n}{n}}{\frac{\sigma}{\sqrt{n}}} - \mu \text{ is approx. standard normally distributed. } \mu \textit{ here is } E(X_n)$$

**Def:** If $X_1,..., X_n$ are independent random variables which are std. norm. dist.ed. Then
$X_1^2 + X_2^2 + ... + X_n^2$ follows the $\chi^2$ distribution with $n$ degrees of freedom. Here each $X$ is a single coordinate in the $n'th$ space. so if n=2, then $X_1, X_2$ describes a point in this space.

**Eg.** Let X follows the $\chi^2$ distribution with n deg of freedom.
Compute $E(X)$ and $Var(X)$
Soln: There are independent standard normally distributed random variables

$X_1,\ldots,X_n$ such that $X = X_1^2 + \ldots + X_n^2$.
$E(X) = E(X_1^2) + E(X_2^2) + \ldots + E(X_n^2) = n$, since $E(X_1^2) = 1$ for all $X$
$Var(X) = Var(X_1^2) + \ldots + Var(X_n^2) = E(X_1^4) - E(X_1^2)^2 + \ldots = 3 - 1^2 + \ldots = 2\,n$

**Lemma:** Let $X_1,\ldots, X_n$ be iids following the normal distribution with mean $\mu$ and variance $\sigma^2$. Then the statistic

$$\frac{\sum_{j=1}^{n}(X_j - \mu)^2}{\sigma^2}$$ follows the $\chi^2$ distribution with n degrees of freedom.

**Pf:** $\frac{X_j - \mu}{\sigma}$ are stand. norm. dist.ed and they are independent. Thus,

$$\sum_j \frac{(X_j - \mu)^2}{\sigma^2}$$ follows the $\chi^2$ distribution with $n$ degrees of freedom.

**Thm:** Let $X_1,\ldots, X_n$ be iids following the normal distribution with mean $\mu$ and variance $\sigma^2$. Then the statistic

$$\frac{\sum_{j=1}^{n}(X_j - (\overline{X}))^2}{\sigma^2}$$, $(\overline{X})$ is the average of some sample. This statistic follows the $\chi^2$ distribution with $n - 1$ deg of freedom.

**Def:** Let X, V be independent random variables following the standard normal and the $\chi^2$ distribution with n deg of freedom resp.. Then

$$\frac{Z}{\sqrt{\frac{V}{n}}}$$ follows the "student" distribution (t distribution) with $n$ deg of freedom.

**Remark:** The student distribution with n deg of freedom is approx. (to two significant figures) the same as the standard normal distribution when n is large (around 30 at least).

**Lemma:** $X_1,\ldots, X_n$ be iids following the normal distribution. We denote $\overline{X} = \frac{1}{n}\cdot(X_1 + \ldots + X_n)$.Then the statistic

$$s^2 = \frac{1}{n-1}\cdot\sum_{j=1}^{n}(X_j - \overline{X})^2$$ is the sample variance. So

$$\frac{\overline{X}}{\sqrt{\frac{s^2}{n}}}$$ follows the student dist. with n-1 deg of freedom.

**Pf:** $\overline{X}\cdot\sqrt{n}$ follows the standard normal dist. and $\sum_{j=1}^{n}(X_j - \overline{X})^2$ follows the $\chi^2$ dist. with $n - 1$ deg of freedom. Then

$$\frac{\overline{X}\cdot\sqrt{n}}{\sqrt{\frac{\sum_{j=1}^{n}\left(X_j-\overline{X}\right)^2}{n-1}}} = \frac{\overline{X}}{\sqrt{\frac{s^2}{n}}}$$ follows the student distribution with $n-1$ deg of freedom.

$evalf\left((x+y)^4\right)$

$$(x+y)^4 \tag{11.1}$$

**Cor:** If $X_1,\ldots,X_n$ are iids following the normal distribution with mean $\mu$ and variance $\sigma^2$. Then

$$\frac{\overline{X}-\mu}{\sqrt{\frac{s^2}{n}}}$$ follows the student distribution with $n-1$ deg of freedom.

$$evalf\left(\frac{92.}{\sqrt{\frac{13^2}{21}}}\right)$$

$$32.43053569 \tag{1}$$

# Estimator 22-10-2019

## Notations

Remark: For each $0 < \alpha < 1$ denote

$P\left(\mathit{Standard\ Normal} > \beta_{\alpha}\right) = \alpha$

$P\left(\chi^2\ \mathit{distribution\ with\ n\ deg\ of\ freedom} > \chi_{n,\alpha}^{2}\right) = \alpha$

$P\left(\mathit{Student\ distribution\ with\ n\ deg\ of\ dfreedom} > t_{n,\alpha}\right) = \alpha$

**Estimator:**

**Def:** Suppose that $X_1,\ldots,X_n$ are iids following a distribution with an unknown parameter $\theta$.

A statistic $f\left(X_1,\ldots,X_n\right)$ is a <u>point estimator</u> of $\theta$ if an experimental result measuring $f\left(X_1,\ldots,X_n\right)$ is regarded as $\theta$.

An experimental result measuring an estimator (a random variable) is called an estimation.

**Eg:** Let $X_1,\ldots,X_n$ be a random sample following a normal distribution, with unkown mean $\mu$. The following statistic

$$\frac{\left(X_1,\ldots,X_n\right)}{n}$$ is an estimator of $\mu$.

So an estimator is an experiment which should represent an unkown parameter, anything can be an estimator, but not all estimators are good or unbiased (Not the same).

**Def:** An estimator $f(X_1,..., X_n)$ of $\theta$ is <u>unbiased</u> if $E(f(X_1,..., X_n)) = \theta$.

**Eg:** Let $X_1,..., X_n$ be a random sample following a normal distribution, with unkown mean $\mu$. The following statistic

$$\frac{(X_1,..., X_n)}{n}$$ is an estimator of $\mu$. Then

$$E\left(\frac{(X_1,..., X_n)}{n}\right) = \frac{1}{n}\cdot\left(E(X_1) + ... + E(X_n)\right) = \frac{1}{n}\cdot(n\cdot\mu) = \mu$$

Therefore the sample average statistic $\frac{(X_1,..., X_n)}{n}$ is an unbiased estimator of $\mu$

**Eg:** Let $X_1,..., X_n$ be a random sample following a normal distribution, with unknown variance $\sigma^2$. The statistic

$$\sum_{j=1}^{n}\frac{(X_j - \overline{X})^2}{n}$$ is an estimator of $\sigma^2$.

**Remember:** $\sum_{j=1}^{n}\frac{(X_j - \overline{X})^2}{\sigma^2}$ follows the $\chi^2$ distribution with $n - 1$ deg of freedom. So that

$$E\left(\sum_{j=1}^{n}\frac{(X_j - \overline{X})^2}{\sigma^2}\right) = n - 1$$

Comparing by changing it yields

$$E\left(\sum_{j=1}^{n}\frac{(X_j - \overline{X})^2}{n}\right) = \frac{n-1}{n}\cdot\sigma^2$$

**Meaning**

$$E\left(\sum_{j=1}^{n}\frac{(X_j - \overline{X})^2}{n}\right) = \frac{n-1}{n}\cdot\sigma^2 \neq \sigma^2$$

Then

$$\sum_{j=1}^{n}\frac{(X_j - \overline{X})^2}{n}$$ is a <u>biased</u> estimator of $\sigma^2$

**However:**

$$E\left(\sum_{j=1}^{n}\frac{(X_j - \overline{X})^2}{n-1}\right) = \sigma^2$$

Then

$$\sum_{j=1}^{n}\frac{(X_j - \overline{X})^2}{n-1}$$ is an ubiased estimator of $\sigma^2$

**Eg:** Let $X_1,..., X_n$ be a random sample following the uniform distribution on $\{1, 2,..., \theta\}$ with a certain unknown $\theta$. Then

$Max(\{X_1, X_2,..., X_n\})$ is an estimator of $\theta$.

$E(Max(\{X_1, X_2,..., X_n\})) = 1\cdot P(\max = 1) + 2\cdot P(\max = 2) + ... + \theta\cdot P(\max = \theta)$ as an ineq.:

$E\left(Max\left(\{X_1, X_2, ..., X_n\}\right)\right) < \theta \cdot P(\max = 1) + \theta \cdot P(\max = 2) + ... + \theta \cdot P(\max = \theta) = \theta$

Then the $Max\left(\{X_1, X_2, ..., X_n\}\right.$ is a biased estimator of $\theta$, but may still be good.

**Eg:** Let $X_1, ..., X_n$ be a random sample following the normal distribution with unknown $\mu$ and known variance $\sigma^2$. Then

$\dfrac{(X_1 + ... + X_n)}{n}$ is an estimator of $\mu$ and follows the normal distribution with mean $\mu$ and variance $\dfrac{\sigma^2}{n}$. Which makes sense, given that the variance gets lower as $n$ gets larger.

$\dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ is standard normally distributed.

In particular $P\left(-1.645 < \dfrac{(\overline{X} - \mu)}{\dfrac{\sigma}{\sqrt{n}}} < 1.1645\right) = 0.9$

$$P\left(-1.645\frac{\sigma}{\sqrt{n}} < (\overline{X} - \mu) < 1.1645\frac{\sigma}{\sqrt{n}}\right) = 0.9$$

$$P\left(\overline{X} - 1.645\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.1645\frac{\sigma}{\sqrt{n}}\right) = 0.9$$

$\mu$ is therefore likely, to be a value in the interval $\overline{X} \pm 1.1645\dfrac{\sigma}{\sqrt{n}}$.

**Def (Interval Estimator):** Let $X_1, ..., X_n$ be a random sample following a distribution with an unknown paramter $\theta$. An <u>interval estimator</u>
is a pair of statistics: $f(X_1, ..., X_n) < g(X_1, ..., X_n)$ , so that $\theta$ lie between an experimental result of $f(X_1, ..., X_n)$ **and** $g(X_1, ..., X_n)$.

**Def (Confidence Interval):** For each $0 < \alpha < 1$. An $\alpha$- confidence interval of an unkown parameter $\theta$ is an interval estimator
$f(X_1, ..., X_n) < g(X_1, ..., X_n)$ such that $P\left(f(X_1, ..., X_n) < \theta < g(X_1, ..., X_n)\right) = \alpha$.

**Lemma:** If $X_1, ..., X_n$ is a random sample following the normal distribution with unkown mean $\mu$ and known variance $\sigma^2$.
Then an $\alpha$-confidence interval of $\mu$ is

$$\overline{X} \pm \frac{z_{(1-\alpha)\cdot 0.5} \cdot \sigma}{\sqrt{n}}$$

Remark In the lemma above, if $X_1, ..., X_n$ are not normally distributed but n is large, then the conculusion on the lemma remains valid.

**Eg:** $X_1, ..., X_n$ is a random sample following the normal distribution with unknown $\mu$ and known variance $\sigma^2$.

Find the sample size $n$ so that the width of a 95 % confidence interval of $\mu$ is less than $\epsilon$.

**Soln:** The width of 95 % confidence interval of $\mu$ is

$$2\cdot\frac{z_{(1-0.95)\cdot 0.5}\cdot\sigma}{\sqrt{n}}<\epsilon$$

$$n>\left(\frac{2\cdot z_{0.025}\cdot\sigma}{\epsilon}\right)^2$$

**Eg:** Let $X_1,\ldots,X_n$ be a the number of success obtained in the 1 *st.*, ...*n th.* Bernoullis trial resp. with unkown probability of succes $p$. Then

$\frac{(X_1+\ldots+X_n)}{n}=\frac{X}{n}$ is a point estimator of $p$. By the Central Limit Thm.

$\frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}$ is approx. standard normally distributed.centr

In particular

$$p\left(-z_{(1-\alpha)\cdot 0.5}<\frac{X-n\cdot p}{\sqrt{n\cdot p(1-p)}}<z_{(1-\alpha)\cdot 0.5}\right)=\alpha$$

# 24-10-2019

$\left|\frac{X-n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}}\right|<z$ in realitty. $p\approx 0.5$, $n\approx 100$, $X\approx 50$, $z\approx 1.6$

$$(X-n\cdot p)^2<z^2\cdot n\cdot p\cdot(1-p)$$

$$(n^2+n\cdot z^2)\cdot p^2+(-2\,n\cdot X-n\cdot z^2)\cdot p+X^2<0$$

$$-\|-<p<\frac{2\,n\cdot X+n\cdot Z^2+\sqrt{(2\cdot n\cdot X+n\cdot z^2)^2-4\cdot(n^2+n\cdot z^2)\cdot X^2}}{2\cdot(n^2+n\cdot z^2)}$$

$$<p<\frac{2\,X+z^2}{2\cdot(n+z^2)}+\frac{\sqrt{4\,n^2\cdot X^2+n^2\cdot z^4+4\,n^2\cdot X\cdot z^2-4\cdot n^2\cdot X^2-4\,n\cdot z^2\cdot X^2}}{2\cdot(n^2+n\cdot z^2)}$$

$$=\frac{2\,X+z^2}{2\cdot(n+z^2)}+\frac{\sqrt{n^2\cdot z^4+4\,n^2\cdot X\cdot z^2-4\,n\cdot z^2\cdot X^2}}{2\cdot(n^2+n\cdot z^2)}=\frac{2\,X+z^2}{2\cdot(n+z^2)}+\frac{\sqrt{z^4+4\cdot X\cdot z^2-\frac{4\,z^2\cdot X^2}{n}}}{2\cdot(n+z^2)}$$

$$\approx\frac{X}{n}+z\frac{\sqrt{\frac{1}{4}\cdot z^2+X-\frac{X^2}{n}}}{n+z^2}\approx\frac{X}{n}+z\frac{\sqrt{\frac{X}{n}\cdot\left(1-\frac{X}{n}\right)}}{\sqrt{n}}=\frac{X}{n}+z\sqrt{\frac{\frac{X}{n}\cdot\left(1-\frac{X}{n}\right)}{n}}$$

Then the inequality is

$$\frac{X}{n}+z\sqrt{\frac{\frac{X}{n}\cdot\left(1-\frac{X}{n}\right)}{n}} < p < \frac{X}{n}+z\sqrt{\frac{\frac{X}{n}\cdot\left(1-\frac{X}{n}\right)}{n}}$$

and the Confidence Interval is given as

$$P\left(\frac{X}{n}+z_{(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{\frac{X}{n}\cdot\left(1-\frac{X}{n}\right)}{n}} < p < \frac{X}{n}+z_{(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{\frac{X}{n}\cdot\left(1-\frac{X}{n}\right)}{n}}\right)=\alpha$$

$$P\left(\overline{X}+z_{(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{\overline{X}\cdot(1-\overline{X})}{n}} < p < \overline{X}+z_{(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{\overline{X}\cdot(1-\overline{X})}{n}}\right)=\alpha$$

**Thm:** If $X_1,\ldots, X_n$ is a random sample of a Bernoulli's trial wth unknown prob of succes p. and $\overline{X}=\frac{(X_1+\ldots+X_n)}{n}$. Then
an $\alpha-$ Confidence Interval of p is

$$\left|\overline{X}-Z\cdot\sqrt{\frac{\overline{X}\cdot(1-\overline{X})}{n}}, \overline{X}+Z\cdot\sqrt{\frac{\overline{X}\cdot(1-\overline{X})}{n}}\right|$$

**Eg:** What is sample size n needed so that an $\alpha$- confidence interval of p above has width $<\epsilon$
**Soln:** We want:

$2\cdot Z_{(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{\overline{X}\cdot(1-\overline{X})}{n}} < \epsilon$ this is guaranteed if

$2\cdot Z_{(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{\frac{1}{4}}{n}}$ or $n > \left(\frac{Z_{(1-\alpha)\cdot 0.5}}{\epsilon}\right)^2$

Note that for each $0\le t\le 1$

$$t(1-t) = -t^2 t - \frac{1}{4}+\frac{1}{4} = -\left(t-\frac{1}{2}\right)^2+\frac{1}{4}\le\frac{1}{4}$$

**Eg:** Let $X_1,\ldots, X_n$ be a random sample following a normal distirbution with known mean $\mu$ and unknown variance $\sigma^2$. Consider

$\sum_{j=1}^{n}\frac{(X_j-\mu)^2}{\sigma^2}$, which follows $\chi^2$- distribution with n deg of freedom.

In particular

$$P\left(\chi^2_{n,\ (1+\alpha)\cdot 0.5} < \sum_{j=1}^{n}\frac{(X_j-\mu)^2}{\sigma^2} < \chi^2_{n,\ (1-\alpha)\cdot 0.5}\right)=\alpha$$

$$P\left(\sum_{j=1}^{n}\frac{(X_j-\mu)^2}{\chi^2_{n,\ (1-\alpha)\cdot 0.5}} < \sigma^2 < \sum_{j=1}^{n}\frac{(X_j-\mu)^2}{\chi^2_{n,\ (1+\alpha)\cdot 0.5}}\right)=\alpha$$

Is then the $\alpha$- CI of $\sigma^2$

**Thm:** If $X_1,\ldots, X_n$ is a random sample following the normal distribution with known mean $\mu$ and unknown variance $\sigma^2$. Then an

$\alpha -$ *Confidence interval* of $\sigma^2$ is

$$\left(\frac{\sum_{j=1}^{n}\left(X_j-\mu\right)^2}{\chi^2_{n,\,(1-\alpha)\cdot 0.5}},\frac{\sum_{j=1}^{n}\left(X_j-\mu\right)^2}{\chi^2_{n,\,(1+\alpha)\cdot 0.5}}\right)$$

$X_1,\ldots,X_n$ *variance.*

Consider

$\sum_{j=1}^{n}\frac{\left(X_j-\overline{X}\right)^2}{\sigma^2}$, which follows the $\chi^2$ distribution with $n-1$ deg of freedom

In particular

$$P\left(\chi^2_{n-1,\,(1+\alpha)\cdot 0.5}<\sum_{j=1}^{n}\frac{\left(X_j-\overline{X}\right)^2}{\sigma^2}<\chi^2_{n-1,\,(1-\alpha)\cdot 0.5}\right)=\alpha$$

$$P\left(\sum_{j=1}^{n}\frac{\left(X_j-\overline{X}\right)^2}{\chi^2_{n-1,\,(1-\alpha)\cdot 0.5}}<\sigma^2<\sum_{j=1}^{n}\frac{\left(X_j-\overline{X}\right)^2}{\chi^2_{n-1,\,(1+\alpha)\cdot 0.5}}\right)=\alpha$$

**Thm:** If $X_1,\ldots,X_n$ is a random sample following the normal distribution with unknown mean $\mu$ and variance $\sigma^2$. Then an

$\alpha -$ *Confidence interval* of $\sigma^2$ is

$$\left(\frac{\sum_{j=1}^{n}\left(X_j-\overline{X}\right)^2}{\chi^2_{n-1,\,(1-\alpha)\cdot 0.5}},\frac{\sum_{j=1}^{n}\left(X_j-\overline{X}\right)^2}{\chi^2_{n-1,\,(1+\alpha)\cdot 0.5}}\right)$$

# 29-10-2019

Let $X_1,\ldots,X_n$ be a random sample following the normal distribution with unknown mean $\mu$ and variance $\sigma^2$ resp.

Consider the statistic

$$\frac{\overline{X}-\mu}{\sqrt{\sum_j \frac{(X_j-\overline{X})^2}{n\cdot(n-1)}}} = \frac{\frac{\frac{\overline{X}-\mu}{\sigma}}{\sqrt{n}}}{\sqrt{\sum_j \frac{(X_j-\overline{X})^2}{\sigma^2}\cdot\frac{1}{\sqrt{n-1}}}} \quad , \sum_j \frac{(X_j-\overline{X})^2}{\sigma^2}$$

follows $\chi^2$ dist. with $n-1$ deg of freedom.
The whole term follows the student distribution with $n-1$ deg of freedom.

In particular $P\left(-t_{n-1,(1-\alpha)\cdot 0.5} < \frac{\overline{X}-\mu}{\sqrt{\sum_j \frac{(X_j-\overline{X})^2}{n\cdot(n-1)}}} < t_{n-1,(1-\alpha)\cdot 0.5}\right) = \alpha$. Isolating $\mu$:

$$P\left(-t_{n-1,(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{S^2}{n}} - \overline{X} < \mu < t_{n-1,(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{S^2}{n}} - \overline{X}\right) = \alpha$$

**Thm:** If $X_1,\ldots,X_n$ is a random sample following the normal distribution with unknown mean $\mu$. Then an $\alpha-$*confidence interval* of $\mu$ is

$$\left(-t_{n-1,(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{S^2}{n}} - \overline{X},\ t_{n-1,(1-\alpha)\cdot 0.5}\cdot\sqrt{\frac{S^2}{n}} - \overline{X}\right), \text{ where } S^2 = \sum_{j=1}^{n}\frac{(X_j-\overline{X})^2}{n-1}$$

## Decisions and determination: Notation and definition

**Scientific Hypothesis:**

**Hypothesis:** Electron exits
**Alternative:** Electrons do not exist.

- Assume that electrons exist, then a cathode ray would be deflected by external E and B field. If the ray does not deflect, then electrons does not exist.
**Point:** Show non existence.
**Result:** Fail

- Assume that electrons exist, then electrons orbit around with calculated energy levels (Based on curent theory).
Experiments are done to measure the energy levels and if it disagrees with current theory. Non-existence of electrons is shown.
**Point:** Show non existence.
**Result:** Fail

**Conclusion:** Non-existence of electrons can not be shown. Existence is not shown, but it's existence can not be denied either.

**Remark:** Scientific hypothesis cannot be proved by experiments. They can only be disproved.

**Def:** In statistic. Hypothesis testing is a decision process.

A hypothesis is set up, which one then plans to disprove. Called the "Null hypothesis" $H_0$.
Then there is an alternative hypothesis. Called $H_1$ or $H_a$, is the statement one plans to except upon succesfull disproving of $H_0$.

**Remark:** The Alternative hypothesis needs not be logical negation of the current $H_0$ in general.

**Remark:** In statistic we usually study a distribution with an unknown paramter $\theta$. $H_0$ is usually
$H_0 : \theta = \theta_0$ or $H_0 : \theta \leq \theta_0$
in any case, the inequality is included as part of $H_0$.

**Eg:** Let X be the daily average temperature at a random day in HK, and $E(X) = \mu$. Test:
$H_0 : \mu = 25\ C$ vs.
$H_1 : \mu \neq 25\ C$
This is an example of a 2-tail test. Logical negation in this case.

**Eg:** Let X,Y be the IQ of a random 4 year old and a random 5 year old resp. and $E(X) = \mu_1$, $E(Y) = \mu_2$. Test
$H_0 : \mu_1 = \mu_2$ vs
$H_1 : \mu_1 < \mu_2$
This is an example of a 1-tail test. Non-negation

**Point 1:** Prove that if x,y are non-negative numbers so that
$x + y = 0$. Then $x = y = 0$ $\left(H_0\right)$.
**Point 2:** Prove that a given coin is unfair $\left(H_0\right)$.

**Pf 1:** Assume that x and y can't both be zero. Then either
$x > 0$ or $y > 0$.
In the former case,
$0 = x + y \geq x > 0$ Contradiction!
In the latter case,
$0 = x + y \geq y > 0$ Contradiction!
Impossible assumption.
Then $x = y = 0$.

**Pf 2:**
Coin:
Assume that this coin is fair ($H_1$), flip it ten times and observe all results to be heads.
$P(10\ consecutive\ heads) = \dfrac{1}{1024}$, which is very unlikely.
So the occurence of this event has a $P = \dfrac{1}{1025}$, which is almost impossible.
Then the coin cannot be fair, and is therefore unfair $\left(H_0\right)$

**Q:** However, what makes it ALMOST impossible, enough to deny $H_1$?

**Remark:** In general, we may design a test as follows:
Flip the coin ten times and let $X$ be the number of heads. We decide to reject $H_0$: the coin is fair when the experiment result of $X$ is in $\{0, 1, 2, 3, 7, 8, 9, 10\}$.

**Remark:** In general, a statistic constists of 2 components:
- a random variable, called the test statistic.
- a subset of $\mathbb{R}$, called the critical region. Then one plans to reject $H_0$ if the experimental result measuring $X$ lies in the critical region. subset?

**Eg:** We plan to test:
$H_0$: the mean height of people in HK $=1.7\ m$
$H_1$ : the mean height of people in HK $\neq 1.7\ m$
Let the test statistic be $\dfrac{(X_1+X_2+\ldots+X_{100})}{100}$
and the critical region is defined as $\mathbb{R}\backslash(1.69, 1.71)$ any real number not in the interval.

## 31-10-2019 Hypothesis Testing Cont.

**Def:** In a hypothesis testing procedure, its type 1 error is the probability that $H_0$ is rejected by the procedure, given that $H_0$ is true.
The type 2 error is the probability that $H_0$ is not rejected by the procedure, given that $H_0$ is false.

**Eg:** Given a coin, where the probability of obatining a head is $p$ when it is flipped. Test $H_0:\ p=\dfrac{1}{2}$ vs $H_1: p\neq\dfrac{1}{2}$
Procedure:
Test Statistic: Let $X_1,\ldots, X_{10}$ be the number of heads obtained in every trial respectively. We make $X_1+X_2+\ldots+X_{10}$ the tet statistic.
Critical Region: $\{0, 1, 2, 3, 7, 8, 9, 10\}$ the critical region.
**Q:** Calculate type 1 error

**Soln: Assume** $H_0$ to be true: so $p=\dfrac{1}{2}$.
$P(H_0\ \textit{is rejected}) = P(0 \leq X_1+X_2+\ldots+X_{10} \leq 3\ \textbf{or}\ 7 \leq X_1+X_2+\ldots+X_{10} \leq 10)$
$=P(\ \textit{the binormial distribution with parameters}\ 10, 0.5) = 1-(10\ C4)\cdot 0.5^4\cdot 0.5^6-(10\ C5)\cdot 0.5^5\cdot 0.5^5-(10\ C6)\cdot 0.5^6\cdot 0.5^4$
$=1-((10\ C4)-(10\ C5)-(10\ C6))\cdot 0.5^{10}$

**Q:** Calculate type 2 error, so that the alternative assumption $p=0.4$ is $H_1$.
**Soln:** $H_1\ p=0.4$
$P(\ X_1+X_2+\ldots+X_{10}=4\ \textbf{or}\ 5\ \textbf{or}\ 6)$

$= P(\textit{binormial distribution with parameter } 10, 0.4) = 4 \textbf{ or } 5 \textbf{ or } 6) = (10\ C4) \cdot 0.4^4 \cdot 0.6^6 - (10\ C5) \cdot 0.4^5 \cdot 0.6^5 - (10\ C6) \cdot 0.4^6 \cdot 0.6^4$

**Remark:** Usually, we make an additional assumption $\theta = \theta_1$ upon rejecting $H_0 : \theta = \theta_0$, so that the type 2 error is computable.

**Eg:** Let $X_1, ..., X_n$ be a random sample following the normal distribution with unknown mean $\mu$ and known variance $\sigma^2$ resp.
We test $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_1$
**Q:** Design a test procedure to test up to a significance level $\alpha$.
**Soln:**
Test statistic: We choose $\frac{X_1, ..., X_n}{n}$ to be the test statistic called $\overline{X}$.
Critical region: Assume that $H_0$ is true, so that $\mu = \mu_0$ Then $\overline{X}$ follows the normal distribution with mean $\mu_0$ and variance $\frac{\sigma^2}{n}$.

In particular $P\left( -Z_{\alpha \cdot 0.5} < \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < Z_{\alpha \cdot 0.5} \right) = 1 - \alpha$.

If we make $\frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ the test statistic and $\mathbb{R} \backslash \left( -Z_{\alpha \cdot 0.5}, Z_{\alpha \cdot 0.5} \right)$ the critical region,
then the type 1 error of such test is $\alpha$.

**Eg:** Let $X_1, ..., X_n$ be a random sample following a certain distribution with unknown mean $\mu$ and known variance $\sigma^2$ resp. and n is large.
Test: $H_0 : \mu = \mu_0$ vs $H_1 = \mu > \mu_0$ up to level $\alpha$.

Type 1 error:
**Soln:** We let $\overline{X}$ be the test statistic, and $(a, \infty)$ the critical region for some $a$.
By the central limit thm, $\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ follows approx the standard normal distribution.

Assume that $H_0$ is true, so that $\mu = \mu_0$ and $\frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ is approx. standard normally distributed.

In particular

$$P\left(Z_\alpha < \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$$

if then $\frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ is the test statistic and $(Z_\alpha, \infty)$ the critical region,

then the type 1 error of such test is $\alpha$.

Type 2 error:
If the additional assumption $\mu = \mu_1$ is made upon rejection of $H_0$.

**Soln:** Assume that $H_0$ is false. In addition $H_1 : \mu = \mu_1$ is assumed.
Then the type 2 error of the test is
$= P(H_0\ \textit{is}\ \textbf{not}\ \textit{rejected}\ \textbf{by}\ \textit{the test})$

$= P\left(\frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < Z_\alpha\right)$ ,but now $\mu \neq \mu_0$, so we don't know how this statistic is distributed. Then

$= P\left(\frac{\overline{X}-\mu_1}{\frac{\sigma}{\sqrt{n}}} + \frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt{n}}} < Z_\alpha\right)$, and by the Central limit thm. $\frac{\overline{X}-\mu_1}{\frac{\sigma}{\sqrt{n}}}$ is approx. standard normally distributed.

$= P\left(\textit{standard normal} < Z_\alpha - \frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$. Look the probability up in a table.

**Eg:** $n = 100$, $\mu_1 - \mu_0 = \frac{1}{10}\sigma$.

$\frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt{n}}} = 1$

We set $\alpha = 0.1$
Table: $Z_\alpha = 1.28$

We set $\beta = 1 - 0.3897 = 0.6103$, by $1.28 - \frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt{n}}} = 1.28 - 1 = 0.28$ and look in table reverse.
Probability that you commit an error of such kind is 61%
Bad test

We set $\alpha = 0.01$
Table:

$Z_\alpha = 2.33$

We set $\beta = 1 - 0.0918 = 0.91$, by $2.33 - 1 = 1.33$ and look in table reverse.
Probability that you commit an error of such kind is 91%
Even worse test.

We set $\alpha = 0.001$
Table: $Z_\alpha = 3.08$

We set $\beta = 1 - 0.0188 = 0.98$, by $3.08 - 1 = 2.08$ and look in table reverse.
Probability that you commit an error of such kind is 98%

**Remark:** One cannot decrease both the type 1 and type 2 errors at the same time, unless the sample size is increased.

# 07-11-2019

In the last Eg that is the sample size needed so that its type 1 and type 2 errors are at most $\alpha$, $\beta$ resp.

**Soln:** We want $P\left( \textit{standard normal} < Z_\alpha - \dfrac{\mu_1 - \mu_0}{\dfrac{\sigma}{\sqrt{n}}} \right) \le \beta$

it is aschieved if $Z_\alpha - \dfrac{\mu_1 - \mu_0}{\dfrac{\sigma}{\sqrt{n}}} \le Z_{1-\beta}$

which happens when $n \ge \left( \dfrac{Z_\alpha - Z_{1-\beta}}{\mu_1 - \mu_0} \sigma \right)^2$

**Eg:** When $\alpha = \beta = 0.9$, $\mu_1 - \mu_0 = \dfrac{1}{10}\sigma$

$n \ge (2 \cdot 1.28 \cdot 10)^2 = 656$

**Eg**: Let p be the unknown probability of success in a Bernoulli's trial. Given samples $X_1,\ldots, X_n$ which are the number of success in the 1st,...,n'th trial resp. and n is large.
test $H_0 : p = p_0$ vs $H_1 := p \ne p_0$ up to a significance level $\alpha$.

**Soln:** Consider $\dfrac{\dfrac{X_1 + \ldots + X_n}{n} - p}{\sqrt{\dfrac{p \cdot (1-p)}{n}}} = \dfrac{\overline{X} - p}{\sqrt{\dfrac{p \cdot (1-p)}{n}}}$ is approx. standard normally distributed by the central limit thm.
Assume that $H_0$ is true.
That is

$$\frac{\overline{X}-p_0}{\sqrt{\frac{p_0\cdot(1-p_0)}{n}}}$$ is approx. standard normally distributed

In particular $P\left(-Z_{\alpha\cdot 0.5} < \frac{\overline{X}-p_0}{\sqrt{\frac{p_0\cdot(1-p_0)}{n}}} < Z_{\alpha\cdot 0.5}\right) = 1-\alpha$

1) We make $\frac{\overline{X}-p_0}{\sqrt{\frac{p_0\cdot(1-p_0)}{n}}}$ the test statistic and

2) $\mathbb{R}\setminus\left(-Z_{\alpha\cdot 0.5}, Z_{\alpha\cdot 0.5}\right)$ the critical region.

Then its type 1 error is $\alpha$.

Eg. Compute the type 2 error of the previous test.
With the additional assumption that $p=p_1$ and that $p_1<p_0$ upon rejection of $H_0$.
**Soln:**Assume that $H_0$ is false, so that $p=p_1$.
Then type 2 error is

$$=P\left(-Z_{\alpha\cdot 0.5} < \frac{\overline{X}-p_0}{\sqrt{\frac{p_0\cdot(1-p_0)}{n}}} < Z_{\alpha\cdot 0.5}\right)$$

What is the distribution? By the previous conclusion that $\frac{\overline{X}-p}{\sqrt{\frac{p\cdot(1-p)}{n}}}$ is approx. standard normally distributed and that $p=p_1$.

Type 2 error is

$$=P\left(-Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{n}} < \overline{X}-p_0 < Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{n}}\right)$$ adding the difference $p_0-p_1$ yields

$$=P\left(-Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{n}}+p_0-p_1 < \overline{X}-p_1 < Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{n}}+p_0-p_1\right)$$

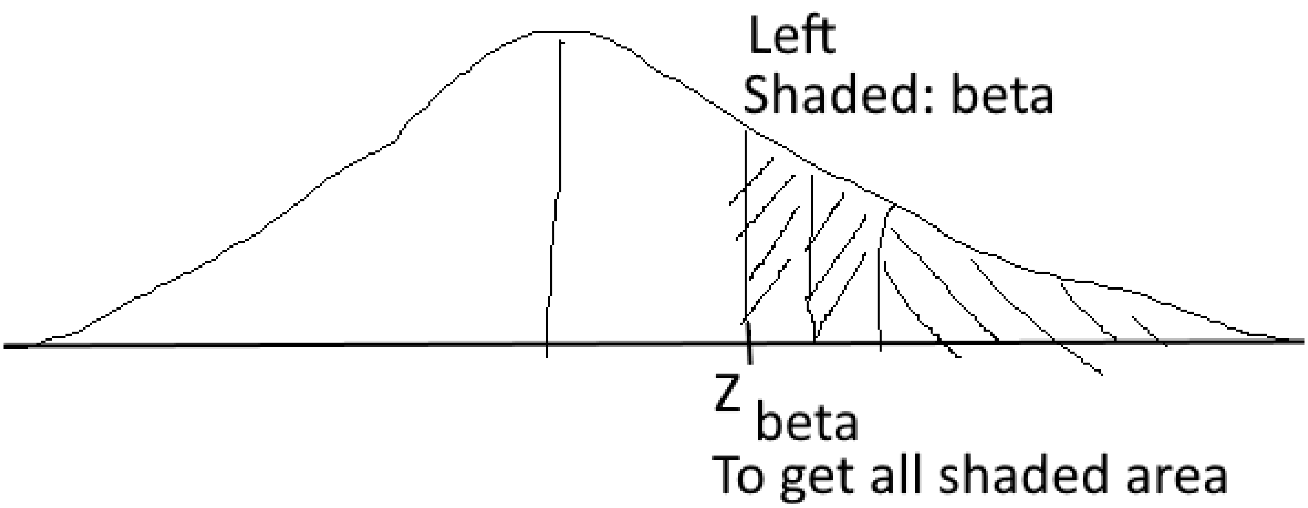
$$=P\left(-Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{p_1\cdot(1-p_1)}}+(p_0-p_1)\cdot\sqrt{\frac{n}{p_1\cdot(1-p_1)}} < \frac{\overline{X}-p_1}{\sqrt{\frac{p_1\cdot(1-p_1)}{n}}} < Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{p_1\cdot(1-p_1)}}+(p_0-p_1)\cdot\sqrt{\frac{n}{p_1\cdot(1-p_1)}}\right)$$

Which is standard normally distributed.

**Eg:** Find the sample size needed so the type 1 and 2 errors of the test are at most α, β resp.

**Soln:** Then we want

$$= P\left( -Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{p_1\cdot(1-p_1)}} + (p_0-p_1)\cdot\sqrt{\frac{n}{p_1\cdot(1-p_1)}} < \frac{\overline{X}-p_1}{\sqrt{\frac{p_1\cdot(1-p_1)}{n}}} < Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{p_1\cdot(1-p_1)}} + (p_0-p_1)\cdot\sqrt{\frac{n}{p_1\cdot(1-p_1)}} \right) \le \beta$$




it is achieved if

$$-Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{p_1\cdot(1-p_1)}} + (p_0-p_1)\cdot\sqrt{\frac{n}{p_1\cdot(1-p_1)}} \ge Z_\beta$$

it happens when

$$n \ge \left(Z_\beta + Z_{\alpha\cdot 0.5}\cdot\sqrt{\frac{p_0\cdot(1-p_0)}{p_1\cdot(1-p_1)}}\right)^2\cdot\frac{p_1\cdot(1-p_1)}{(p_0-p_1)^2}$$

**Eg:**
$\alpha = \beta = 0.1,\ p_0 = 0.5,\ p_1 = 0.4.$
$Z_\beta = 1.28$ and $Z_{\alpha\cdot 0.5} = 1.645$

$$\left(1.28 + 1.645\cdot\sqrt{\frac{0.5\cdot(1-0.5)}{0.4\cdot(1-0.4)}}\right)^2\cdot\frac{0.4\cdot(1-0.4)}{(0.5-0.4)^2}$$

$$210.1251370 \tag{16.1}$$

Then the sample size needed is 211

**Eg:** Let X be normally distributed with known mean $\mu$ and unknown variance $\sigma^2$. Given samples $X_1, \ldots, X_n$ following the distribution of $X$. Test

$H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$ up to a level $\alpha$.

**Soln:** Consider $\dfrac{\sum_j (X_j - \mu)^2}{\sigma^2}$ follows the $\chi^2$ distribution with n deg. of freedom.

Assume that $H_0$ is true.

So that $\dfrac{\sum_j (X_j - \mu)^2}{\sigma^2}$ follows the $\chi^2$ distribution with n deg. of freedom.

In particular

$$P\left(\frac{\sum_j (X_j - \mu)^2}{\sigma^2} > \chi^2_{n,\,\alpha}\right) = \alpha$$

We make $\dfrac{\sum_j (X_j - \mu)^2}{\sigma^2}$ the test statistic and

$\left(\chi^2_{n,\,\alpha}, \infty\right)$ the critical region.

Then the type 1 error of this test is $\alpha$.

**Eg:** Let X be normally distributed with unknown mean $\mu$ and variance $\sigma^2$ and $X_1 \ldots X_n$ is a sample of it. Test

$H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ up to a level $\alpha$.

**Soln:** Consider $\dfrac{\sum_j (X_j - \overline{X})^2}{\sigma^2}$ which follows the $\chi^2$ distribution with $n - 1$ deg. of freedom.

Assume that $H_0$ is true so that

$\dfrac{\sum_j (X_j - \overline{X})^2}{\sigma_0^2}$ follows the $\chi^2$ distribution with $n - 1$ deg. of freedom.

$$P\left(\chi^2_{n-1,\,1-(\alpha \cdot 0.5)} < \frac{\sum_j (X_j - \overline{X})^2}{\sigma_0^2} < \chi^2_{n-1,\,\alpha \cdot 0.5}\right) = 1 - \alpha$$

Test statistic: $\dfrac{\sum_j \left(X_j - \overline{X}\right)^2}{\sigma_0^2}$

Critical region: $\left(\chi^2_{n-1,\, \alpha\cdot 0.5}, \infty\right)$ U(and) $\left(0, \chi^2_{n-1,\, 1-(\alpha\cdot 0.5)}\right)$

## 19-11-2019

Let X be normally distributed with unknown mean and unknown variance. Given a random sample $X_1,\ldots, X_n$ distributed the same way as X

Test $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$ up to a significance $\alpha$.

Consider $\dfrac{\overline{X} - \mu}{\sum_{j=1}^{n} \dfrac{\left(X_j - \overline{X}\right)^2}{n-1}}$ , $\dfrac{\left(X_j - \overline{X}\right)^2}{n-1}$ is the sample variance. Used instead of $\sigma^2$ since it is unknown.

Assume that $H_0$ is true, so that X is normally distributed with mean $\mu = \mu_0$.

$$= \frac{\dfrac{\overline{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}}{\left(\sum_{j=1}^{n} \dfrac{\left(X_j - \overline{X}\right)^2}{(n-1)\cdot\sigma^2}\right)^{\frac{1}{2}}}$$

, Denominator is standard normally distributed and nominator is $\chi^2$ - distributed with $n-1$ deg of freedom.

This follows the student distribution with $n-1$ deg of freedom.

In particular

$$P\left(\frac{\overline{X} - \mu}{\dfrac{s}{\sqrt{n}}} > t_{n-1,\, \alpha}\right) = \alpha$$

So we make $\dfrac{\overline{X} - \mu}{\dfrac{s}{\sqrt{n}}}$ the test statistic and

$\left(t_{n-1,\, \alpha}, +\infty\right)$ the critical region.

The type 1 error is $\alpha$.

**Eg:** Let X be the average temperature at a random day in November.

Assume that X is normally distributed with mean $\mu$.

Test $H_0$: $\mu = 21.8$ vs $H_1 : \mu > 21.8$. So a 1-tail test.

Use R, find $t$ value and determine whether it falls within the critical region.