BITS Pilani
Pilani Campus

Introduction to Statistics

innovate achieve lead

BITS Pilani
Pilani Campus

Course No: DSECL ZC413
Course Title: ISM
WEBINAR 2

Topics - Webinar

- Bayes' and Naïve Bayes Classifiers
- Discrete Random Variable
- Continuous Random Variable
- Probability distribution and density function
- Mean Variance and Standard deviation
- Joint Probability function
- Binomial Distribution

Topics - Webinar

- Poisson Distribution
- Normal Distribution
- Uniform Distribution

Conditional Probability

$\frac{{}^{50}C_1}{} = \frac{50}{} = \frac{1}{}$.

Example: Suppose three coins tossed simultaneously. Let E and F be two events which are defined as follows:

E- At least Two heads appear.

F-First Coin shows tail.

a) Find $P(E)$. b) If it is given that the first coin shows tail then what is $P(E)$?

Solution Let S be a sample space.

$\therefore S = \{HHH, HHT, THH, HTH, TTH, HTT, THT, TTT\}$

$n(S) = 8$,

$E = \{HHH, HHT, THH, HTH\}$

$F = \{TTT, THH, THT, TTH\}$

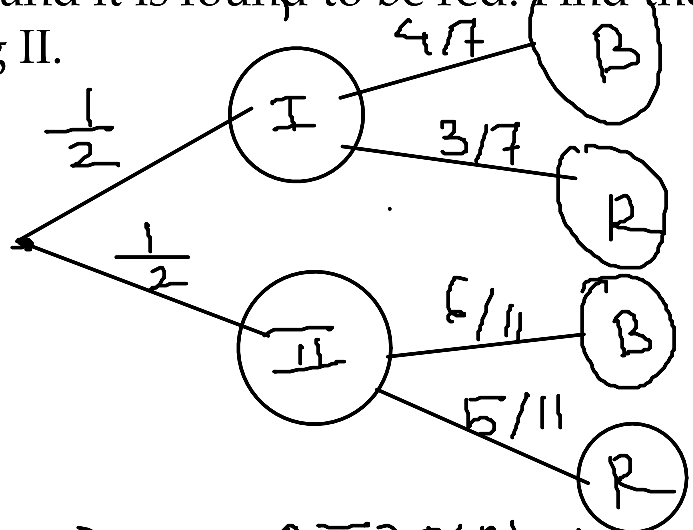
(a) $P(E) = \frac{n(E)}{n(S)} = \frac{4}{8} = \frac{1}{2}$.

(b) $P(E|F) = \frac{n(E \cap F)}{n(F)} = \frac{P(E \cap F)}{P(F)} = \frac{1}{4}$.

Bayes' Theorem

Example: Bag I contains 3 red and 4 black balls and while another Bag II contains 5 red and 6 black balls. One ball is drawn at random from one of the bags and it is found to be red. Find the probability that it was drawn from Bag II.

Solution

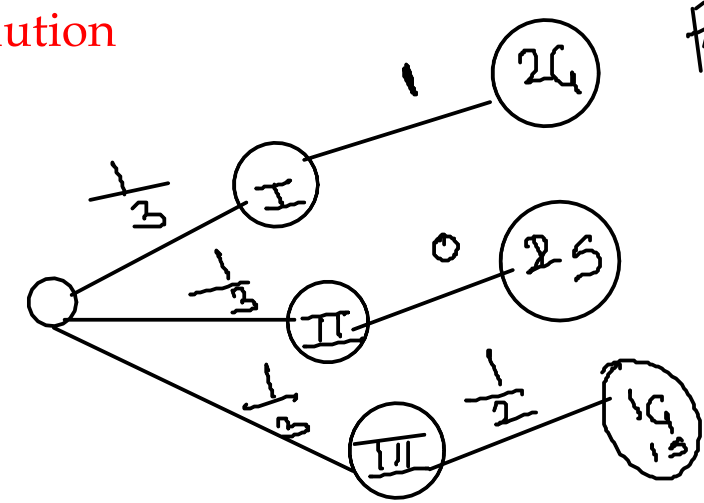


$$P(II \mid R) = \frac{P(II)\,P(R \mid II)}{P(R)} = \frac{\frac{1}{2} \times \frac{5}{11}}{\frac{1}{2} \times \frac{3}{7} + \frac{1}{2} \times \frac{5}{11}}$$

Bayes' Theorem

Example: Given three identical boxes I,II and III, each containing two coins . In box I both coins are gold coins, in box II, both are silver coins and in the box III there is one gold and one silver coin. A person chooses a box at random and takes out a coin. If the coin is gold, what is the probability that the other coin in the box is also of gold.

Solution



Prob of selecting gold coin

$$= P(G) = \frac{1}{3}\times 1 + \frac{1}{3}\times 0 + \frac{1}{3}\times\frac{1}{2}$$

$$= \frac{1}{3} + 0 + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}.$$

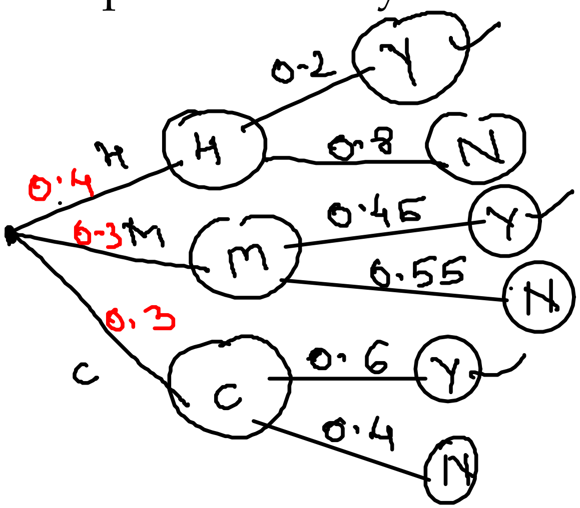
From getting other coin also gold, we have to select first coin from bag I.

$$\therefore \text{Required prob} = P(I|G) = \frac{P(I)\,P(G|I)}{P(G)} = \frac{\frac{1}{3}\times 1}{1/2} = \frac{2}{3}.$$

Bayes' Theorem

Example : From past record it is known that the chance of hot temperature , mild and cool temperature on a particular day is 0.4,0.3 and 0.3 respectively. The chance of playing golf is depending on temperature of that day. If temperature is hot then chance of playing golf is 0.2, for mild temperature chance is 0.45 and for cool temperature chance is 0.6. If the team plays golf on a particular day then find the probability that the day was hot.

Solution



Prob of playing golf game

$$P(Y) = (0.4)\times(0.2) + (0.3)(0.45) + (0.3)(0.6) = 0.395$$

$$0.08 + 0.135 + 0.18$$

$$P(H|Y) = \frac{P(H)\,P(Y|H)}{P(Y)} = \frac{0.4 \times 0.2}{0.395} = \frac{0.08}{0.395}$$

Naive Bayes Classifiers

Consider a fictional dataset that describes the weather conditions for playing a game of golf. Given the weather conditions, each tuple classifies the conditions as fit(“Yes”) or unfit(“No”) for playing golf.

	Outlook	Temperature	Humidity	Windy	Play Golf
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Naive Bayes Classifiers

If Given that the weather conditions are “Rainy outlook”, “Temperature is hot”, “high humidity” and “no wind” then predict that the golf match will be played or not under such condition.

Solution

$$P(Y / R, H, HH, F)$$
$$= P(R/Y)\, P(H/Y)\, P(HH/Y)\, P(F/Y) \cdot P(Y)$$

$$P(N / R, H, HH, F) = P(R/N)\, P(H/N) \cdots P(N)$$

Naive Bayes Classifiers

$P(H/Y) = \frac{P(H \cap Y)}{9} = \frac{2}{9}$

$E = \{2,4,6\}$

$F = \{1,3,5\}$

$= \frac{P(R \cap Y)}{9}$

$E \cap F = \phi.$

Outlook

	Yes	No	P(yes)	P(no)
Sunny	2	3	2/9 ✓	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9 $\frac{2}{9}$	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9 ✓	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

$\{HH, HT, TH, TT\}$

$E = \{HH\}$

$F = \{TT\}$

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

$E \cap F = \phi.$

Naive Bayes Classifiers

Suppose today = (Sunny, Hot, Normal, False)

So, probability of playing golf is given by:

$$P(Yes|today) = \frac{P(SunnyOutlook|Yes)P(HotTemperature|Yes)P(NormalHumidity|Yes)P(NoWind|Yes)P(Yes)}{P(today)}$$

and probability to not play golf is given by:

$$P(No|today) = \frac{P(SunnyOutlook|No)P(HotTemperature|No)P(NormalHumidity|No)P(NoWind|No)P(No)}{P(today)}$$

Since, P(today) is common in both probabilities, we can ignore P(today) and find proportional probabilities as:

$$P(Yes|today) \propto \frac{2}{9}.\frac{2}{9}.\frac{6}{9}.\frac{6}{9}.\frac{9}{14} \approx 0.0141$$

$P(A)+P(A^c)=1$

And

$$P(No|today) \propto \frac{3}{5}.\frac{2}{5}.\frac{1}{5}.\frac{2}{5}.\frac{5}{14} \approx 0.0068$$

Now, since

$$P(Yes|today) + P(No|today) = 1$$

These numbers can be converted Into a probability by making the sum equal to 1 (normalization):

$$P(Yes|today) = \frac{0.0141}{0.0141+0.0068} = 0.67$$ and $$P(No|today) = \frac{0.0068}{0.0141+0.0068} = 0.33$$

Since $P(Yes|today) > P(No|today)$.

So, prediction that golf would be played is ‘Yes’.

Discrete Random Variable & their Probability Distribution

1. Suppose a random variable X takes on the values $-4,2,3,7$ with respective probabilities $\frac{k+2}{10}, \frac{2k-3}{10}, \frac{3k-4}{10}, \frac{k+1}{10}$. Find the distribution and expected value of X.

$\Rightarrow$ Here $f(-4) = \frac{k+2}{10}$, $f(2) = \frac{2k-3}{10}$, $f(3) = \frac{3k-4}{10}$, $f(7) = \frac{k+1}{10}$

since $f(x)$ is prob mass function

$\therefore \sum f(x) = 1$

$$\frac{k+2+2k-3+3k-4+k+1}{10} = 1$$

$$\frac{7k-4}{10} = 1 \Rightarrow k = 2.$$

$\therefore$ Distribution is $f(-4) = \frac{4}{10}$, $f(2) = \frac{1}{10}$, $f(3) = \frac{2}{10}$, $f(7) = \frac{3}{10}$

The expected value of $X = \sum x f(x) = -4 \times \frac{4}{10} + 2 \cdot \frac{1}{10} + 3 \times \frac{2}{10} + 7 \times \frac{3}{10}$

$$= \frac{-16+2+6+21}{10} = \frac{13}{10} = 1.3$$

Discrete Random Variable & their Probability Distribution

2. Let X be a random variable with following distribution

x	1	3	4	5
$f(x)$	0.4	0.1	0.2	0.3

Find i) the mean ii) variance iii) standard deviation iv) $E(3X+2)$ v) $E(X^2)$ vi) $E(2^x)$

$\Rightarrow$ i) the mean $= \sum x f(x) = 1\times\frac{4}{10} + 3\times\frac{1}{10} + 4\times\frac{2}{10} + 5\times\frac{3}{10}$

$= \frac{4+3+8+15}{10} = \frac{30}{10} = 3.$

ii) variance $= \sigma^2 = E(x^2) - (E(x))^2$ —— 1

$E(x^2) = \sum x^2 f(x) = 1\times\frac{4}{10} + 9\times\frac{1}{10} + 16\times\frac{2}{10} + 25\times\frac{3}{10}$

$= \frac{4+9+32+75}{10} = 12$

$\therefore \sigma^2 = 12 - (3)^2 = 12 - 9 = 3$

iii) $S.D = \sqrt{3}$.

iv) $E(3X+2) = 3E(X)+2 = 3\times 3+2 = 11.$

v) $E(X^2) = 12$

vi) $E(2^x) = \sum 2^x f(x) = 2^1\times\frac{4}{10} + 2^3\times\frac{1}{10} + 2^4\times\frac{2}{10} + 2^5\times\frac{3}{10}$

$= \frac{8}{10} + \frac{8}{10} + \frac{32}{10} + \frac{96}{10} = \frac{144}{10}.$

innovate achieve lead

3. A fair coin is tossed until a head or five tails occur. Find the expected number of tosses of the coin.

$\Longrightarrow$ Sample Space $= \{H, TH, TTH, TTTH, TTTTH, TTTTT\}$

X be a random which counts number of tosses

$\therefore X$ assigns $\{1, 2, 3, 4, 5\}$ values to sample space.

The pmf $f(x)$ is, $f(1) = \frac{1}{2}$, $f(2) = \frac{1}{4}$, $f(3) = \frac{1}{8}$, $f(4) = \frac{1}{16}$

$$f(5) = \frac{1}{32} + \frac{1}{32} = \frac{1}{16}.$$

$\therefore$ Expected number of tosses $= \sum x f(x)$

$$= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + 5 \cdot \frac{1}{16}$$

$$= \frac{8 + 8 + 6 + 4 + 5}{16} = \frac{31}{16} \approx 2$$

4. A box contains 8 light bulbs of which 3 are defective. A bulb is selected from the box and tested. If it is defective, another bulb is selected and tested, until a non-defective bulb is chosen. Find the expected number of bulbs chosen.

$\Rightarrow$ If we write D for defective and N for non-defective then sample space S has four elements.

$$S = \{N, DN, DDN, DDDN\}$$

Let x be a random variable which represent the number of bulbs chosen.

$\therefore$ x assigns $\{1, 2, 3, 4\}$ values to sample points.

and pmf is, $f(1) = \frac{5}{8}$, $f(2) = \frac{3}{8} \frac{5}{7} = \frac{15}{56}$, $f(3) = \frac{3}{8} \frac{2}{7} \times \frac{5}{6} = \frac{5}{56}$

$f(4) = \frac{3}{8} \frac{2}{7} \frac{1}{6} \frac{5}{5} = \frac{1}{56}$.

$\therefore E(x) = \sum x f(x) = 1 \cdot \frac{5}{8} + 2 \cdot \frac{15}{56} + 3 \cdot \frac{5}{56} + 4 \cdot \frac{1}{56} = \frac{3}{2}$.

Discrete Random Variable & their Probability Distribution

5. A game is played by two person A and B as follows:

"A throws two dice. If the sum is 7 he wins \$3 from B. If the sum is 8 he loses \$2 and if sum is 3 he loses \$4, otherwise no money changes hand. Is this a fair game?"

$\Rightarrow$ Let x be a random variable which represent the gain by person A.

$\therefore$ X assigns $\{0, 3, -2, -4\}$ to sample space.

Let f be prob distribution function.

$\therefore f(3) = P\{\text{getting sum 7 after throwing two dice}\}$

$= P\{(1,6), (6,1), (2,5), (5,2), (3,4), (4,3)\} = \frac{6}{36}$.

$f(-2) = P\{\text{getting sum 8}\}$

$= P\{(2,6), (6,2), (4,4), (3,5), (5,3)\} = \frac{5}{36}$.

$f(-4) = P\{\text{getting sum 3}\}$

$= P\{(1,2), (2,1)\} = \frac{2}{36}$

$f(0) = P\{\text{getting sum other than 7, 8, and 3}\} = \frac{23}{36}$.

$\therefore E(X) = 0 \times \frac{23}{36} + 3 \times \frac{6}{36} - 2 \times \frac{5}{36} - 4 \times \frac{2}{36} = 0$

$\therefore$ Game is fair.

Discrete Random Variable & their Probability Distribution

6. A linear array EMPLOYEE has n elements. Suppose NAME appears randomly in the array, and there is a linear search to find the location K of NAME, that is, to find K such that EMPLOYEE[K]=NAME. Let X denote the number of comparisons in the linear search. Find the expected value of *X*.

$\Rightarrow$ Let X denote number of comparisons. Since NAME appears randomly in any position in the array with the same prob $\frac{1}{n}$. We have $X = 1, 2, 3, \ldots, n$ each probability with $1/n$.

Hence $E(x) = 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + \cdots + n \cdot \frac{1}{n} = \frac{n+1}{2}$

Continuous Random Variable & their Probability Distribution

7. Let X be random variable with probability density function $f(x) = \frac{1}{2}e^{-|x|}$ for all $x \in \mathbb{R}$. If $Y = X^2$ then find cumulative distribution function of Y.

$$\Rightarrow \quad F(Y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le x \le \sqrt{y})$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} e^{-|x|} dx$$

But $e^{-|x|}$ is an even function

$$\therefore F(Y) = 2\int_0^{\sqrt{y}} \frac{1}{2} e^{-x} dx = \int_0^{\sqrt{y}} e^{-x} dx = -e^{-x} \Big/_0^{\sqrt{y}}$$

$$= 1 - e^{-\sqrt{y}}$$

$$\therefore F(Y) = \begin{cases} 1 - e^{-\sqrt{y}} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Continuous Random Variable & their Probability Distribution

8. Let X be random variable with probability density function $f(x) = \begin{cases} x^2\left(2x+\frac{3}{2}\right) & 0 < x \leq 1 \\ 0 & otherwise \end{cases}$ If $y = \frac{2}{x} + 3$ then find $var(y)$.

$$\Rightarrow \quad Var(y) = 4\, Var\left(\frac{1}{x}\right) = 4\left\{E(1/x^2) - \left(E(1/x)\right)^2\right\}$$

$$E(1/x^2) = \int_{-\infty}^{\infty} \frac{1}{x^2} f(x)\,dx = \int_0^1 \frac{1}{\cancel{x^2}} \cancel{x^2}\left(2x+\frac{3}{2}\right)dx$$

$$= \frac{2x^2}{2} + \frac{3}{2}x \Big|_0^1 = \frac{5}{2}.$$

$$E(1/x) = \int_0^1 \frac{1}{x} x^2\left(2x+\frac{3}{2}\right)dx$$

$$= \int_0^1 \left(2x^2 + \frac{3}{2}x\right)dx = \frac{2x^3}{3} + \frac{3}{2}\frac{x^2}{2}\Big|_0^1$$

$$= \frac{2}{3} + \frac{3}{4}\Big|_0^1 = \frac{14}{12} = \frac{7}{6}$$

$$\therefore Var(Y) = 4\left\{\frac{5}{2} - \left(\frac{7}{6}\right)^2\right\} = 4\left\{\frac{5}{2} - \frac{49}{36}\right\}.$$

Continuous Random Variable & their Probability Distribution

9. The distribution of amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous random variable X with probability density function

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$. What is the cumulative distributive function of sales for any X? How do you use this to find the probability that $X < 0.25$? What about the probability that X is greater than 0.75? What about $p(0.25 < X < 0.75)$.

$\Rightarrow$ Let $F(x)$ be cdf, $\therefore F(x) = p(X \le x) = \int_{-\infty}^{x} f(x)\,dx$

If $x < 0$, then $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{x} 0\,dx = 0.$

If $0 \le x < 1$ then $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{x} \frac{3}{2}(1-x^2)\,dx$

$= \frac{3}{2}\left(x - \frac{x^3}{3} \Big|_0^x\right) = \frac{3}{2}\left(x - \frac{x^3}{3}\right)$ —— (1)

If $x \geq 1$ then $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} \frac{3}{2}(1-x^2)\,dx + \int_{1}^{x} 0\,dx$

$= 1.$

(b) $p(X < 0.25) = F(0.25) = \frac{3}{2}\left(0.25 - \frac{(0.25)^3}{3}\right)$ from (i)

(c) $p(X > 0.75) = \int_{0.75}^{\infty} f(x)\,dx = \int_{3/4}^{1} \frac{3}{2}(1-x^2)\,dx$

$$= \frac{3}{2}\left(x - \frac{x^3}{3} \Big|_{3/4}^{1}\right)$$

$$= \frac{3}{2}\left\{1 - \frac{1}{3} - \frac{3}{4} + \frac{(3/4)^3}{3}\right\}$$

$$= \frac{3}{2}\left\{\frac{2}{3} - \frac{3}{4} + \frac{27}{64\times 3}\right\}.$$

(d) $P(0.25 < X < 0.75) = \int_{0.25}^{0.75} f(x)\,dx = \int_{1/4}^{3/4} \frac{3}{2}(1-x^2)\,dx$

$$= \frac{3}{2}\left(x - \frac{x^3}{3} \Big|_{1/4}^{3/4}\right) = \frac{3}{2}\left\{\frac{3}{4} - \frac{1}{3}\left(\frac{3}{4}\right)^3 - \frac{1}{4} + \frac{1}{3}\left(\frac{1}{4}\right)^3\right\}$$

$$= \frac{3}{2}\left\{\frac{3}{4} - \frac{9}{64} - \frac{1}{4} + \frac{1}{64\times 3}\right\}$$

Continuous Random Variable & their Probability Distribution

10. If cumulative distribution function is $F(x) = \begin{cases} 0 & x \le 0 \\ (1-e^{-x})^2 & x > 0 \end{cases}$. Find probability density function and $p(1 < X < 2)$.

$\Rightarrow$ Let $p(x)$ be density function of x.

$$\therefore f(x) = \frac{d}{dx} F(x) = \begin{cases} 0 & x \le 0 \\ 2(1-e^{-x})e^{-x} & x > 0 \end{cases}$$

$$= \begin{cases} 0 & x \le 0 \\ 2(e^{-x} - e^{-2x}) & x > 0 \end{cases}$$

$$\therefore p(1 < x < 2) = 2\int_1^2 (e^{-x} - e^{-2x})\,dx = 2\left\{ -e^{-x} + \frac{e^{-2x}}{2} \Big|_1^2 \right\}$$

$$= 2\left\{ -e^{-2} + \frac{e^{-4}}{2} + e^{-1} - \frac{e^{-2}}{2} \right\}$$

$$= 2\left\{ -\frac{3e^{-2}}{2} + \frac{e^{-4}}{2} + e^{-1} \right\}$$

Joint Probability Distribution

11. Let X denote the number of times a photocopy machine will malfunction: 0, 1, 2, or 3 times, on any given month. Let Y denote the number of times a technician is called on an emergency call. The joint p.m.f. $p(x, y)$ is presented in the table below:

X

Y

	0	1	2	3
0	0.15	0.30	0.05	0
1	0.05	0.15	0.05	0.05
2	0	0.05	0.10	0.05

Find a) $P(Y > X)$ b) Marginal probability function of X and Y. c) $P(X \leq 2, Y \geq 1)$.

(a) $p(Y>X) = p(0,1) + p(0,2) + p(1,2)$

$= 0.05 + 0 + 0.05 = 0.1$

(b) Marginal prob function of $X = f_1(x) = \sum_y f(x,y)$

$\therefore f_1(0) = f(0,0) + f(0,1) + f(0,2) = 0.15 + 0.05 + 0 = 0.2$.

$f_1(1) = f(1,0) + f(1,1) + f(1,2) = 0.30 + 0.15 + 0.05 = 0.5$

$f_1(2) = f(2,0) + f(2,1) + f(2,2) = 0.05 + 0.05 + 0.10 = 0.2$

$f_1(3) = f(3,0) + f(3,1) + f(3,2) = 0 + 0.05 + 0.05 = 0.1$

Marginal prob function of $Y = f_2(y) = \sum_x f(x,y)$.

$f_2(0) = f(0,0) + f(1,0) + f(2,0) + f(3,0) = 0.15 + 0.30 + 0.05 + 0 = 0.5$

$f_2(1) = f(0,1) + f(1,1) + f(2,1) + f(3,1) = 0.05 + 0.15 + 0.05 + 0.05 = 0.3$

$f_2(2) = f(0,2) + f(1,2) + f(2,2) + f(3,2) = 0 + 0.05 + 0.10 + 0.05 = 0.2$

ⓒ $P(X \leq 2, Y \geq 1) = f(0,1) + f(0,2) + f(1,1) + f(1,2) + f(2,1) + f(2,2)$

$= 0.30 + 0.05 + 0.15 + 0.05 + 0.05 + 0.10$

$= 0.7$.

Joint Probability Distribution

12. Let the joint density function for (X, Y) be $f(x,y) = \begin{cases} \frac{c(x+y)}{3} & 0 < x < 2, 0 < y < 1 \\ 0 & ohrtwise \end{cases}$

Find a) the constant c b) $P(X > Y)$ c) Marginal density function of X and Y.

d) Are X and Y are independent?

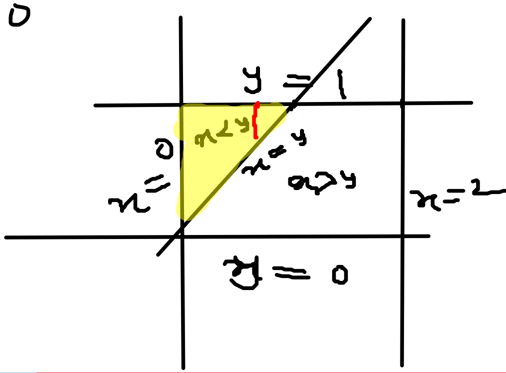
(a) since $f(x,y)$ is a density function

$\therefore \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx = 1$; $\int_0^2\int_0^1 \frac{c(x+y)}{3}\,dy\,dx = 1$

$\therefore \frac{c}{3}\int_0^2 \left. xy + \frac{y^2}{2} \right|_0^1 dx = 1 \Rightarrow \frac{c}{3}\int_0^2 \left(x + \frac{1}{2}\right) dx = 1$

$\Rightarrow \frac{c}{3}\left\{ \left. \frac{x^2}{2} + \frac{x}{2} \right|_0^2 \right\} = 1$

$\Rightarrow \frac{c}{3}\left\{ \frac{4}{2} + \frac{2}{2} \right\} = 1 \Rightarrow c = 1.$



(b) $P(X > Y) = 1 - P(X \leq Y)$

$= 1 - \int_0^1\int_x^1 \frac{x+y}{3}\,dy\,dx$

$$= 1 - \frac{1}{3}\int_0^1 xy + \frac{y^2}{2}\Big|_x^1 dx = 1 - \frac{1}{3}\int_0^1 x + \frac{1}{2} - \frac{3x^2}{2}\, dx$$

$$= 1 - \frac{1}{3}\left\{ \frac{x^2}{2} + \frac{x}{2} - \frac{3x^3}{6}\Big|_0^1 \right\}$$

$$= 1 - \frac{1}{3}\left\{ \frac{1}{2} + \cancel{\frac{1}{2}} - \cancel{\frac{1}{2}} \right\}$$

$$= 1 - \frac{1}{6} = \frac{5}{6}$$

marginal density function of X $= f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$

$$= \int_0^1 \frac{x+y}{3}\,dy$$

$$= \frac{1}{3}\left\{ xy + \frac{y^2}{2}\Big|_0^1 \right\} = \begin{cases} \frac{1}{3}\left(x+\frac{1}{2}\right) & 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

marginal density function of Y $= f_2(y)$

$$= \int_{-\infty}^{\infty} f(x,y)\,dx = \frac{1}{3}\int_0^2 x+y\,dx$$

$$= \frac{1}{3}\left\{ \frac{x^2}{2} + xy\Big|_0^2 \right\} = \begin{cases} \frac{1}{3}(2+2y) & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(d) consider $f_1(x)\cdot f_2(y) = \frac{1}{3}\frac{(2x+1)}{2}\cdot\frac{(2+2y)}{3}$

$$\ne \frac{x+y}{3} = f(x,y)$$

$\therefore$ X and Y are not independent.

Joint Probability Distribution

13. Dick and Jane have agreed to meet for lunch between noon (0:00 p.m.) and 1:00 p.m. Denote Jane's arrival time by X, Dick's by Y, and suppose X and Y are independent with probability density functions $f_x(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$, $f_y(y) = \begin{cases} 2y & 0 \le y \le 1 \\ 0 & otherwise \end{cases}$

Find the probability that Jane arrives before Dick.

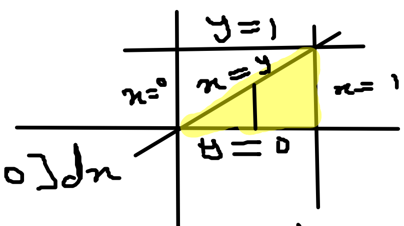
$\Rightarrow$ Here it is given that X and Y are independent

$$\therefore f(x,y) = f_x(x)\, f_y(y)$$

$$= \begin{cases} 6x^2y & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$P(\text{Jane arrives before Dick}) = P(X<Y)$$

$$= \int_0^1 \int_0^x 6x^2 y \, dy\, dx$$



$$= \int_0^1 \frac{6x^2y^2}{2}\Big|_0^x dx = 3\int_0^1 x^2[x^2-0]dx$$

$$= 3\int_0^1 x^4 dx = 3\left\{ \frac{x^5}{5}\Big|_0^1 \right\} = \frac{3}{5}.$$

Binomial Distribution

14. If from six to seven in the evening one telephone line in every five is engaged in a conversation: what is the probability that when 10 telephone numbers are chosen at random, only two are in use?

$\Longrightarrow$ Let x be a random variables which represant the number of telephone lines in use.

This is Binomally distributed random variable because telephone line either in use or not in use.

Let the telephone line in use will be success.

$\therefore$ Prob (success) $= \frac{1}{5}$

$n = 10$, $p(x=2) = \binom{10}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^8$.

Binomial Distribution

15. It has been determined that 5% of drivers checked at a road stop show traces of alcohol and 10% of drivers checked do not wear seat belts. In addition, it has been observed that the two infractions are independent from one another. If an officer stops five drivers at random:

a) Calculate the probability that exactly three of the drivers have committed any one of the two offenses.

b) Calculate the probability that at least one of the drivers checked has committed at least one of the two offenses.

$\Rightarrow$ Let A be an event that driver traces of alcohol.
B be an event that driver do not wear seat belt.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = (0.05) + 0.01 - (0.05)(0.01)$
$= 0.145.$

Here success is driver commited any one the two offenses.

$\therefore$ $P(\text{success}) = 0.145$.

(a). $n = 5$, $p = 0.145$, $q = 0.855$

$p(x=3) = {}^5C_3 (0.145)^3 (0.855)^2 = 0.0223$

(b) $p(x \geq 1) = 1 - p(x=0) = 1 - {}^5C_0 (0.145)^0 (0.855)^5$.

$= 0.543$.

Binomial Distribution

16. A student takes an 18-question multiple choice exam, with four choices per question. Suppose one of the choice is obviously incorrect and the students make an "educated" guess of the remaining choices. Find the expected number E of correct answers and the standard deviation σ.

$\Longrightarrow$ Let x be a random variable which count the number of correct answers.

Here success is giving correct answer to the quetion.

$p(\text{success}) = \frac{1}{3}$ {because one of the choice is obviously incorrect}.

As x is binomally distributed

$\therefore$ Expected value of $x = E(x) = np = 18 \times \frac{1}{3} = 6$

and $S.D = \sqrt{npq} = \sqrt{18 \times \frac{1}{3} \times \frac{2}{3}} = \sqrt{4} = 2.$

Poisson Distribution

17. Suppose 300 misprints are distributed randomly throughout a book of 500 pages. Find the probability P that a given page contains: (a) Exactly 2 misprints (b) 2 or more misprints.

$\Rightarrow$ We view the number of misprints on one page as the number of success in a seq of Bernoullis trials.

Here $n = 300$, prob of getting a mistake on a given page $= \frac{1}{500}$.

$\therefore \lambda = np = \frac{300}{500} = 0.6$.

As number of misprints are more, so we can solve this problem by Poisson distribution.

(a) $p(x=2) = \frac{(0.6)^2 e^{-0.6}}{2!} = 0.0988$.

(b) $p(x \geq 2) = 1 - \{p(0) + p(1)\}$

$= 1 - \left\{ \frac{(0.6)^0 e^{-0.6}}{0!} + \frac{(0.6)^1 e^{-0.6}}{1!} \right\}$

$= 0.122$.

Poisson Distribution

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

18. The number of industrial injuries per working week in a particular factory is known to follow a Poisson distribution with mean 0.5. Find the probability that (a) in a particular week there will be: (i) less than 2 accidents, (ii) more than 2 accidents; (b) in a three-week period there will be no accidents.

$\Rightarrow$ Here $\lambda = 0.5$

Let x be a random variable which count number of accidents in industry which is Poissonaly distr

① $p(x<2) = p(x=0) + p(x=1)$

$$= \frac{e^{-0.5}(0.5)^{0}}{0!} + \frac{e^{-0.5}(0.5)^{1}}{1!}$$

② $p(x>2) = 1 - \{p(0) + p(1) + p(2)\}$

$$= 1 - \left\{\frac{e^{-0.5}(0.5)^{0}}{0!} + \frac{e^{-0.5}(0.5)^{1}}{1!} + \frac{e^{-0.5}(0.5)^{2}}{2!}\right\}$$

③ In a three week period, the avg of number of accidents $= 3\times\lambda$

$$\therefore p(x=0) = \frac{e^{-(3\times\lambda)}(1.5)^{0}}{0!} = (e^{-\lambda})^{3} = (e^{-0.5})^{3}.$$

Normal Distribution

19. Suppose the weights of 2000 male students are normally distributed with $\mu = 155lb$ and standard deviation $\sigma = 20\ lb$. Find the number of students with weights: a) not more than 100lb b) between 150 and 175 lb (inclusive) c) greater than or equal to 200lb.

$\Longrightarrow$ Let x be random variables which measure wt of students is normally distributed with

$\mu = 155, \ \sigma = 20.$

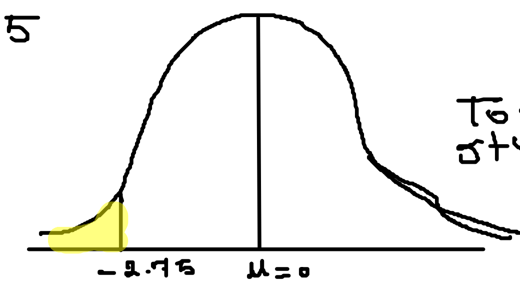
(a) $P(X \le 100) = ?$

Let $z = \dfrac{100 - 155}{20} = \dfrac{-55}{20} = -2.75$

$\therefore P(X \le 100) = P(z \le -2.75)$

$= 0.5 - \phi(2.75)$

$= 0.5 - 0.4970 = 0.03$



Total students

$\therefore$ num of students ht not more than 100 lb $= 0.03 \times 2000$
$= 60.$

(b) $P(150 \le X \le 175)$

std unit of $150 = \dfrac{150 - 155}{20} = \dfrac{-5}{20} = -0.25.$

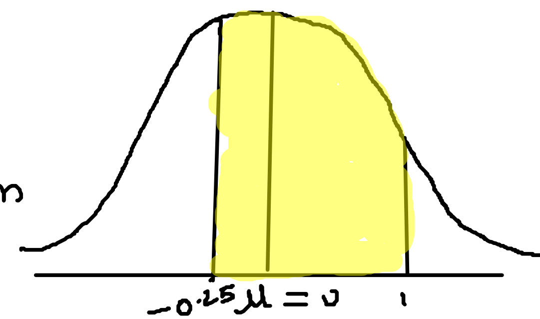
$$\text{std unit of } 175 = \frac{175-155}{20} = 1.$$

$$\therefore\ p(150 \le X \le 175) = p(-0.25 \le Z \le 1)$$

$$= \phi(0.25) + \phi(1)$$

$$= 0.0987 + 0.3413$$

$$= 0.44.$$



$\therefore$ number of students with ht between 150 and 175 lb $= 0.44 \times 2000$

$$= 880.$$

(c) $p(X \geq 200) = ?$

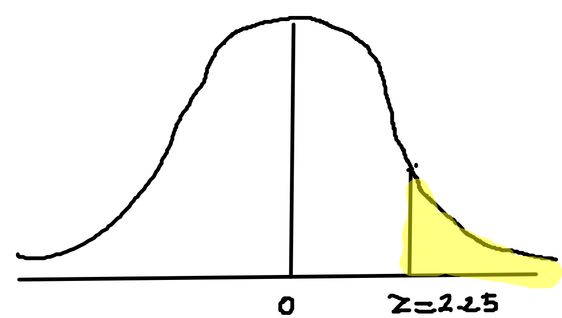
$$\text{std unit of } 200 = \frac{200-155}{20} = \frac{45}{20} = 2.25.$$

$$p(X \geq 200) = p(Z \geq 2.25)$$

$$= 0.5 - \phi(2.25)$$

$$= 0.5 - 0.4878$$

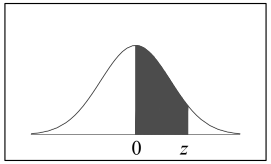
$$= 0.0122.$$



$\therefore$ Number of students with ht more than 200 lb $= 2000 \times 0.0122$

$$= 24.4 \approx 24$$

Standard Normal Distribution Table



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995

Normal Distribution

20. A fair die is tossed 720 times. Find the probability that the face 6 will occur: Between 100 and 125 times b) more than 130 times c)less than 110 times.

$\Rightarrow$ This is problem of Binomial distribution with large number of trials; and

mean $= \mu = np = 720 \times \frac{1}{6} = 120$, $\sigma = \sqrt{npq} = 10$

Here np is also large, so we can solve this problem by using normal distribution.

(a) $BP(100 \le x \le 125) = NP(99.5 \le x \le 125.5)$.

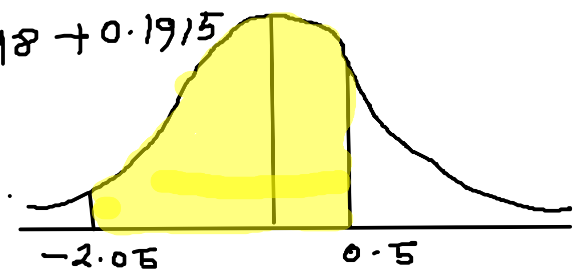
$\therefore$ std unit of $99.5 = \frac{99.5 - 120}{10} = -2.05$

std unit of $125 = \frac{125 - 120}{10} = 0.5$

$\therefore NP(99.5 \le x \le 125.5) = NP(-2.05 \le z \le 0.5)$

$= \phi(2.05) + \phi(0.5) = 0.4798 + 0.1915$

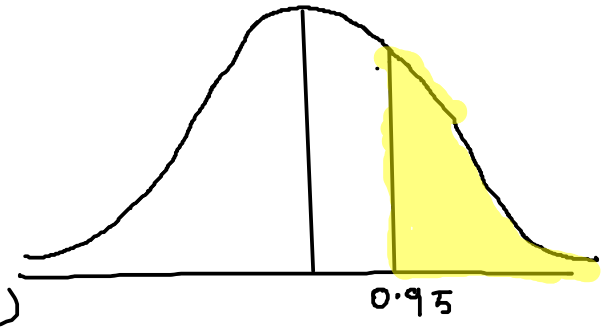
$= 0.6713.$



(b) $BP(X \geq 130) = NP(X \geq 129.5)$

$$\text{Std unit of } 129.5 = \frac{129.5 - 120}{10}$$

$$= 0.95$$



$$NP(X \geq 129.5) = NP(Z \geq 0.95)$$

$$= 0.5 - \phi(0.95)$$

$$= 0.5 - 0.3289 = 0.1711.$$

(c) $BP(X \leq 110) = NP(X \leq 110.5)$

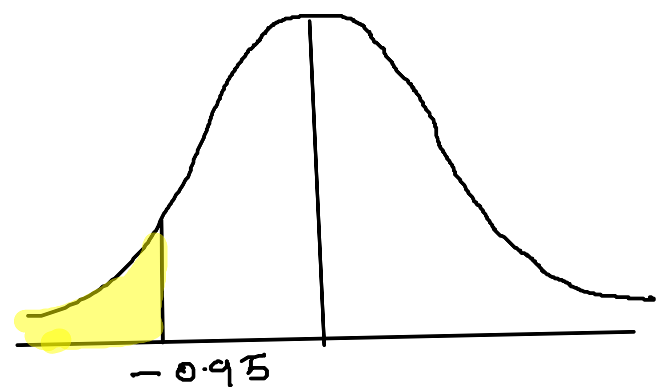
$$\text{Std unit of } 110.5 = \frac{110.5 - 120}{10} = -0.95.$$

$$NP(X \leq 110.5) = NP(Z \leq -0.95)$$

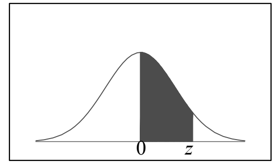
$$= 0.5 - \phi(0.95)$$

$$= 0.5 - 0.3289$$

$$= 0.1711.$$



Standard Normal Distribution Table



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997

Gilles Cazelais. Typeset with LaTeX on April 20, 2006.

Thank you

IMP Note to Self

STOPREC