BSCMA1004

STATISTICS II NOTES

WEEK 4 NOTES

IITM B.S Degree

PREPARED BY

Vehaan Handa, IIT Madras

Statistics II - Week IV Notes

★ Continuous Random Variables

A drv X takes values in alphabet $\mathcal{X}$. $|\mathcal{X}|$ is growing very large, is unwieldy for calculations.

Eg. Meteorite weight, Binomial (n,p)

Meteorite Data:

① Preprocessing: Take $\log_2 \rightarrow [0.01, 60000000]$

$$\downarrow$$

$$[-6.6, 25.8]$$

We still have 45000+ data

② Move from individual values to intervals.
Divide $[-6.6, 25.8] \approx 100$ intervals.
$[-6.6, -6.3], [-6.3, -6] \ldots [25.5, 25.8]$
Count no. of values falling inside each interval.

★ (Cumulative) Distribution Function [CDF]

The CDF of a rv X, $F_X(x)$ is a fn. from $\mathbb{R}$ to $[0,1]$

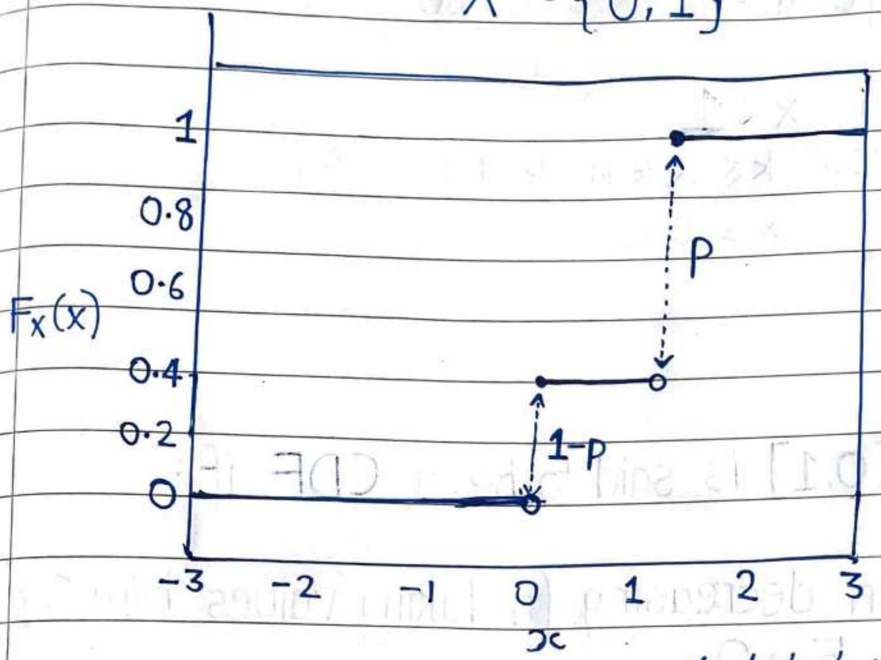
$$F_X(x) = P(X \leq x)$$

$F_X(b) - F_X(a) = P(a < X \leq b)$
F_X: non decreasing, non negative values
As $x \rightarrow -\infty$, $F_X \rightarrow 0$
As $x \rightarrow \infty$, $F_X \rightarrow 1$

DATE

Eg. Bernoulli RV

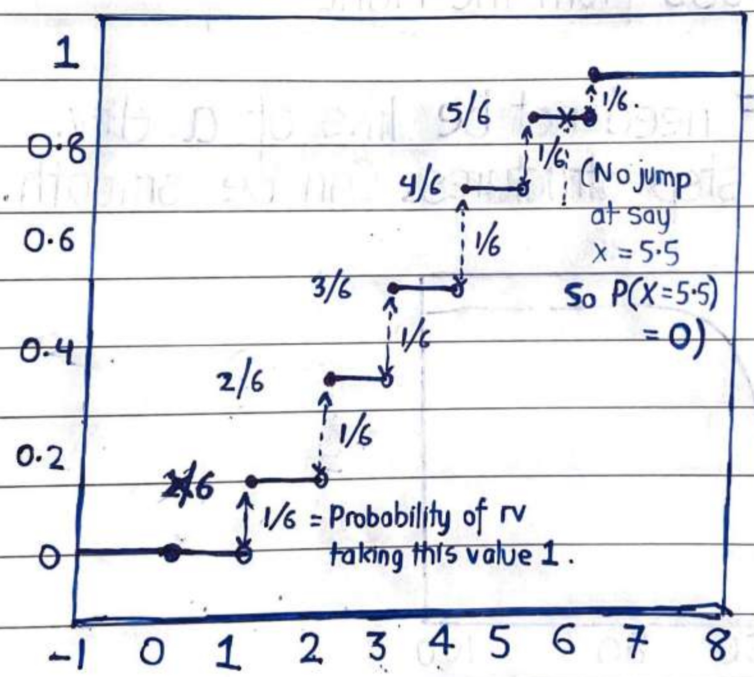
$$X \sim \{\overset{1-p}{0}, \overset{p}{1}\}$$



$$F_X(x) = \begin{cases} 0 & x<0 \\ 1-p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Eg. Die throw

$$X \sim \{\overset{1/6}{1}, \overset{1/6}{2}, \overset{1/6}{3}, \overset{1/6}{4}, \overset{1/6}{5}, \overset{1/6}{6}\}$$



DATE

★ Computing Probability of Intervals using CDF

$X \sim \text{Uniform}\{1, 2, \ldots\ldots, 100\}$

$$F_X(x) = \begin{cases} 0 & x < 1 \\ k/100 & k \leq x < k+1,\ k = 1, 2, \ldots, 99. \\ 1 & x \geq 100. \end{cases}$$

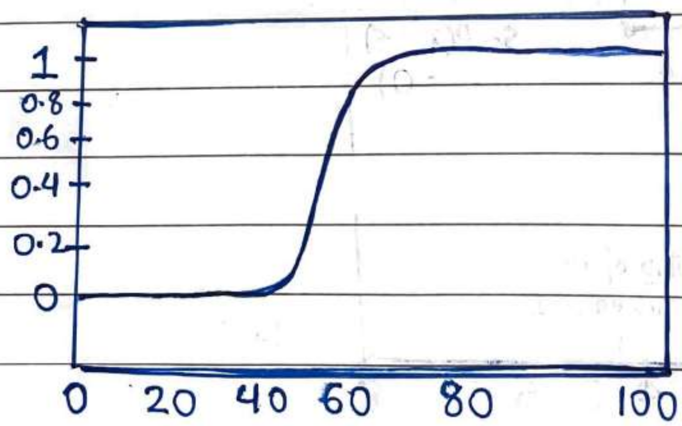
★ CDF

A fn. $F : \mathbb{R} \rightarrow [0,1]$ is said to be a CDF if:

1. F is a non decreasing fn taking values b/w 0 and 1.
2. As $x \rightarrow -\infty$, $F \rightarrow 0$
3. As $x \rightarrow \infty$, $F \rightarrow 1$
4. F is continuous from the right

A general CDF need not be like of a drv.

- No need for step structures. Can be smooth.



★ Binomial using continuous CDF

$X \sim \text{Binomial}(100, 0.6)$

$$F_X(k) = \sum_{j=0}^{k} {}^{100}C_j \, (0.6)^j (0.4)^{n-j} \qquad F(x) = \frac{1}{1 + \exp\left(\frac{-1.65451(x-60)}{\sqrt{24}}\right)}$$

np

$(np(1-p))$

- $P(40 < X \leq 50) = 0.0271$ $F(50) - F(40) = 0.0318$
- $P(50 < X \leq 60) = 0.5108$ $F(60) - F(50) = 0.4670$
- $P(60 < X \leq 70) = 0.4473$ $F(70) - F(60) = 0.4670$
- $P(70 < X \leq 80) = 0.0148$ $F(80) - F(70) = 0.0318$.

★ CDF & Random Variables

Given a valid CDF $F(x)$, $\exists$ rv X taking values in $\mathbb{R}$ such that

$$P(X \leq x) = F(x)$$

- If $F(x)$ rises from F_1 to F_2 at x_1, $P(X = x_1) = F_2 - F_1$
- If $F(x)$ is continuous at x_0, $P(X = x_0) = 0$

* $P(1.99 < X < 2.01) = F(2.01) - F(1.99) = 0.002$ (say)
- Value with finite precision taken with +ve probability

* $P(1.9999999 < X < 2.0000001) = 0.0000002$

* $P(X = 2.000000\ldots) = 0$
- Cannot take value with infinite precision when fn is cont.

DATE

★ Continuous Random Variables

A random var. X with CDF $F_X(x)$ is said to be a continuous RV if $F_X(x)$ is continuous at every x.

- CDF has no jumps/steps
- $P(X=x) = 0$ for all x.
- $P(a<X\leqslant b) = F(b) - F(a)$

$\because P(X=a) = 0 = P(X=b)$

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

Many drvs are well approximated by crv's.

★ Probability Density Function (PDF)

A crv X with CDF $F_X(x)$ is said to ~~half~~ have a PDF $f_X(x)$; if $\forall x_0$,

$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)\,dx$$

- CDF is the integral of PDF
- Value of PDF around $f_X(x_0) \equiv X$ takes a value around x_0.
  - Higher the PDF, higher this chance
- ∴ The PDF can be >1, if range of values is shorter.

A func. $f : \mathbb{R} \to \mathbb{R}$ is said to be a density function if:

① $f(x) \geq 0$

② $\int_{-\infty}^{\infty} f(x)\,dx = 1$

③ $f(x)$ is piecewise continuous

- Given a density fn f, there is a crv X with PDF as f.

- Support of the random variable X with PDF f_X is $\text{supp}(X) = \{x : f_X(x) > 0\}$

- $\text{supp}(X)$ contains intervals in which X can fall with +ve probability. $\int_{\text{supp}(X)} f_X(x)\,dx = 1$.

- Remember $P(X=x) = 0$ for a CRV.

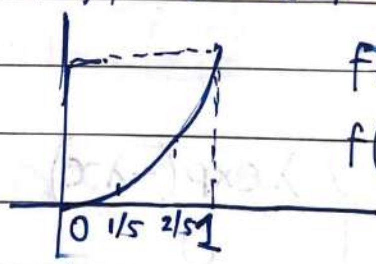
For rv X with PDF f_X, event A is a subset of real line and prob. is $P(A) = \int_A f(x)\,dx$

Q. Consider $f(x) = \begin{cases} 3x^2 & 0<x<1 \\ 0 & \text{otherwise} \end{cases}$

Show f is a density fn. Consider rv X with density f.
Find $P(X=1/5)$, $P(X=2/5)$, $P(X \in [1/5-\epsilon, 1/5+\epsilon])$.

$P(X=1/5) = 0$

$P(X=2/5) = 0$



$f(1/5) = 3/25$

$f(2/5) = 12/25$

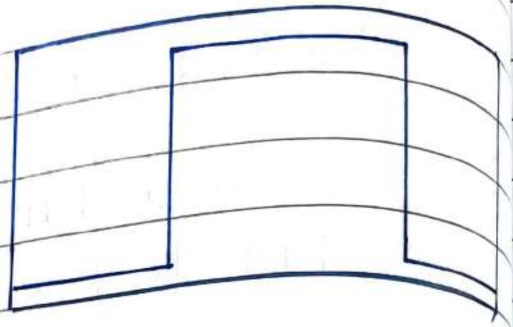
$$P(1/5-\epsilon, 1/5+\epsilon) = \int_{1/5-\epsilon}^{1/5+\epsilon} 3x^2\,dx = x^3 \Big|_{1/5-\epsilon}^{1/5+\epsilon}$$

$$= \left(\frac{1}{5}+\epsilon\right)^3 - \left(\frac{1}{5}-\epsilon\right)^3 = \frac{6}{25}\epsilon + \cancel{2\epsilon^3} \ll \epsilon$$

★ Common Distributions

$X \sim \text{Uniform}[a,b]$

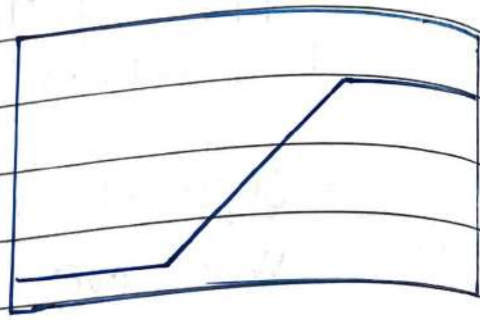
PDF



- PDF

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

CDF



- CDF

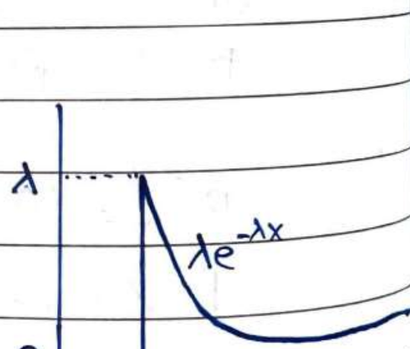
$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a<x<b \\ 1 & x \geq b \end{cases}$$

Suppose $X \sim \text{Uniform}[-10,10]$. Find $P(-3 \leq X \leq 2)$ $\nearrow 5/20$, $P(5<|X|<7)$ $\swarrow 4/20$, $P(1-\epsilon < X < 1+\epsilon)$ $\searrow 2\epsilon/20$ for $x_0 \in [-9,9]$, $P(X>7 \mid X>3)$ $\downarrow 3/7$

$$f_X(x) = \begin{cases} 1/20 & -10<x<10 \\ 0 & \text{otherwise} \end{cases}$$

$X \sim \text{Exp}(\lambda)$

- PDF



$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & x>0 \\ 0 & \text{otherwise} \end{cases}$$

- CDF

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1-\exp(-\lambda x) & x>0 \end{cases}$$

DATE

$X \sim Exp(2)$

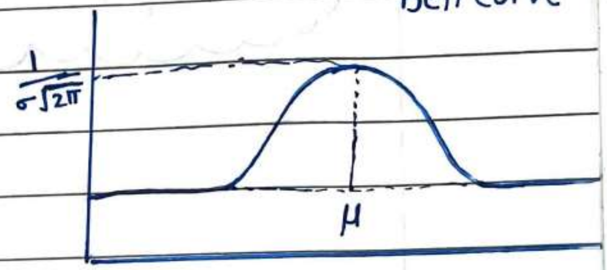
$$P(1-\varepsilon < X \leq 1+\varepsilon) = e^{-2(1-\varepsilon)} - e^{-2(1+\varepsilon)}$$

$P(X > s+t \mid X > s) = e^{-t}$ → Unique for exp. distribution. "memoryless"

Normal Distribution / Gaussian Distribution

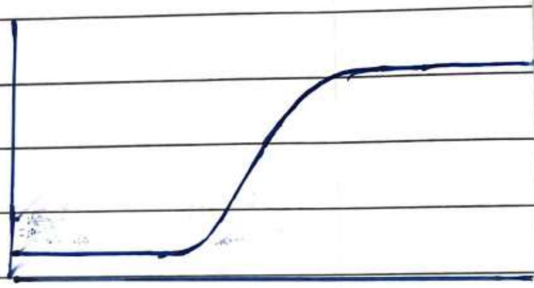
$X \sim \text{Normal}(\mu, \sigma^2)$ $\mu \to$ any real num, $\sigma \to$ +ve real num

"Bell Curve"



- PDF $\text{supp}(X) = \mathbb{R}$

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



- CDF

no closed form

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du.$$

Standard Normal : Normal(0,1)

- CDF of $X \sim \text{Normal}(\mu, \sigma^2)$ does not have closed form.
- Standardisation : If $X \sim \text{Normal}(\mu, \sigma^2)$ then :→ $(X-\mu/\sigma) \sim \text{Normal}(0,1)$

$Z \sim \text{Normal}(0,1)$ PDF : $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$

CDF : $F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)\, du$

Normal tables : Tabulation of $F_Z(z)$

So convert prob. to a standard normal. Use normal tables.

DATE

$X \sim \text{Normal}(\overset{\mu}{3}, \overset{\sigma^2}{1})$

$Z = X - 3 \sim \text{Normal}(0,1)$ $\quad X = Z + 3$

$P(5 < X < 7) \leftrightarrow 5 < Z + 3 < 7 \Rightarrow 2 < Z < 4$

$= P(2 < Z < 4) = F_Z(4) - F_Z(2)$

$P(X > \mu) = P(X < \mu) = 1/2$