Example:

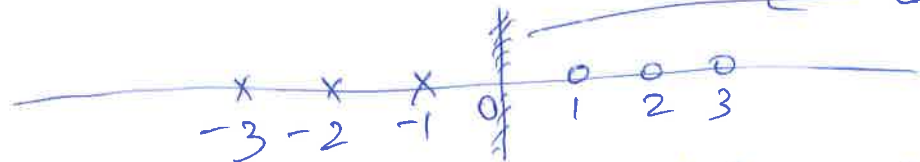
Two classes and we observe some data:

Training data

class 1	1, 2, 3
class 2	-1, -2, -3

← feature values

← decision threshold



Let us model class 1 data using the Gaussian (Normal) prob. density function (p.d.f.)

$$f_X(x \mid \text{class}1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}(x-\mu_1)^2\right)$$

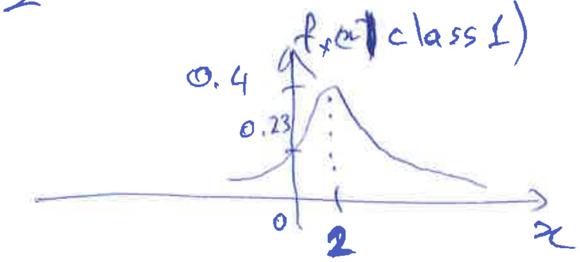
# of observations

Mean: $\hat{\mu}_1 = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{3}(1+2+3) = \frac{6}{3} = 2$

Variance: $\hat{\sigma}_1^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\hat{\mu}_1)^2 = \frac{1}{2}\left((1-2)^2+(2-2)^2+(3-2)^2\right) = 1$

Standard deviation $\hat{\sigma}_1 = 1$

$$f_X(x|1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$$



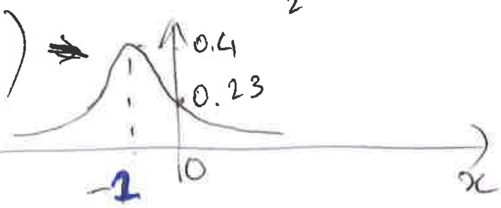
Class 2 training: $f_X(x \mid \text{class } 2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2\sigma_2^2}(x-\mu_2)^2\right)$

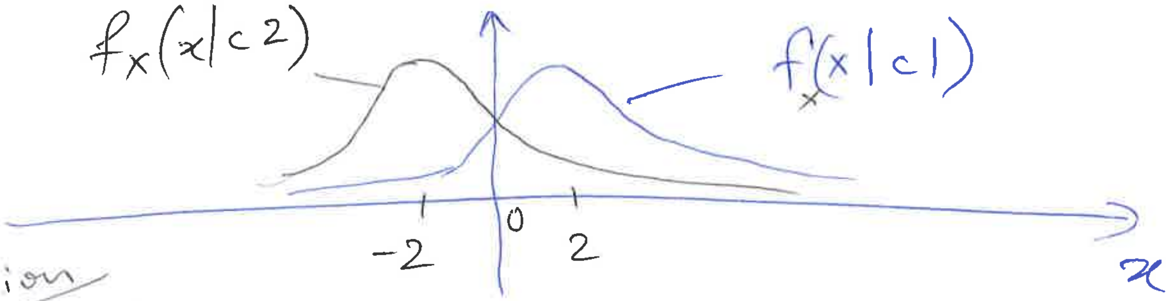
Mean of class 2: $\hat{\mu}_2 = \frac{1}{3}(-1-2-3) = -2$

Variance of " " : $\hat{\sigma}_2^2 = \frac{1}{2}\left((-1+2)^2+(-2+2)^2+(-3+2)^2\right) = 1$

Std. dev. " " : $\hat{\sigma}_2 = 1$

$f_X(x \mid c2)$ →





The decision threshold is 0.

$$f_X(x|c1) > f_X(x|c2) \implies \text{class 1}$$

$$f_X(x|c2) > f_X(x|c1) \implies \text{class 2}$$

Testing Procedure

So If the new observation $x > 0 \implies$ class 1

$x < 0 \implies$ class 2.

Let us assume that I observe $x = 4$

$$f_X(4|c1) > f_X(4|c2) \implies \text{class 1}.$$

Total Probability Theorem for p.d.f.'s.

$$f_X(x) = f_X(x|c_1)\,P(c1) + f_X(x|c_2)\,P(c_2)$$

where Prior probabilities $P(c1)$ is the probability of class 1,

$P(c2)$ " " " " " 2

We can estimate the prior probabilities from our observations: We have 6 observations. class 1 has 3, class 2 has 3.

$$\hat{P}(c1) = \frac{3}{6}, \quad \hat{P}(c_2) = \frac{3}{6} = \frac{1}{2}.$$

Random variables X & Y are independent

$$f_{XY}(x, y) = f_X(x)\,f_Y(y)$$

Cumulative Distribution Function (CDF)

Jan 22/1

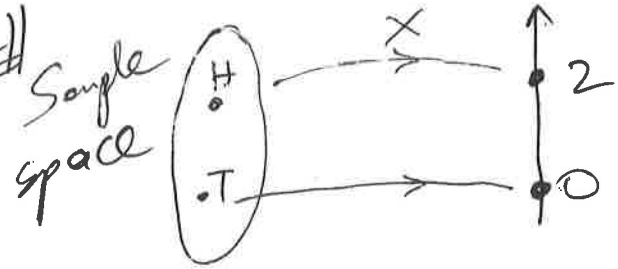
CDF: $F_X(x) = P_{rob}(X \leq x)$

r.v. — dummy variable

$F_X(x) = P(\text{Random variable} \leq x)$ $\in \mathbb{R}$.

r.v.

Ex



r.v. is defined

$X(H) = 2$

$X(T) = 0$.

$P(H) \triangleq P(T) \triangleq 1/2$

$F_X(-5) = P(X \leq -5) = 0$

$F_X(x) = P(X \leq x) = 0$ when $x < 0$

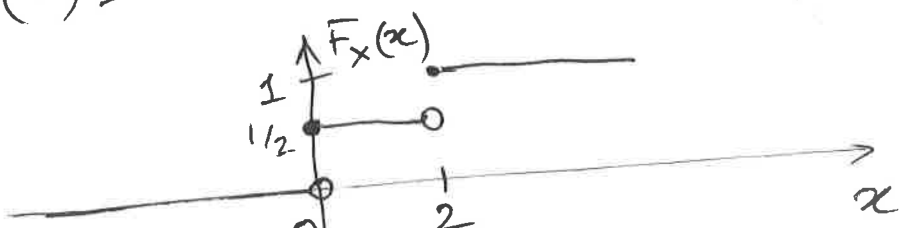
$F_X(0) = P(X \leq 0) = P(X(T)=0) = 1/2$ at $x = 0$

$= P(X = 0)$

when $F_X(x) = P(X \leq x) = P(X=0) = 1/2$ when $0 \leq x < 2$

$F_X(x) = P(X \leq 2) = P(X=0) + P(X=2) = 1$ when $x = 2$

$F_X(x) = 1$ when $x \geq 2$.



Probability Density function (PDF)

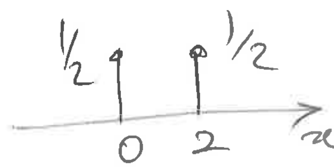
$$f_X(x) = \frac{d F_X(x)}{dx}$$

Ex cont'd $f_X(x) = \frac{1}{2}\delta(x) + \frac{1}{2}\delta(x-2)$

Jan 22/2

Probability Mass Function (PMF) $\equiv$ PDF for discrete r.v.

$$P_X(x) = \begin{cases} 1/2, & X = 2 \\ 1/2, & X = 0 \\ 0, & \text{otherwise} \end{cases}$$



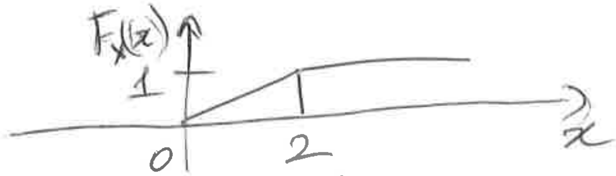
Ex Continuous random variable. example.

I pick a number between 0 and 2 at random without any bias.

$P(X=0.5) = 0$ became length $(0.5) = 0$

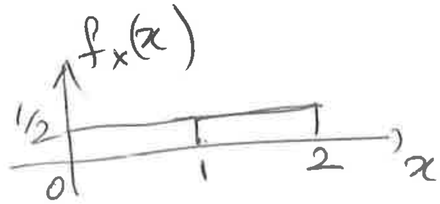
$P(0.1 \le X \le 0.2) = \dfrac{\text{length } [0.1, 0.2]}{\text{length } [0, 2]} = \dfrac{0.1}{2} = 0.05$

C.D.F. $F_X(x) = \begin{cases} 0, & 0 < x \\ \frac{1}{2}x, & 0 \le x \le 2, \quad \dfrac{\text{length } [0, x]}{\text{length } [0,2]} = \dfrac{x}{2} \\ 1 & x > 2 \end{cases}$



p.d.f. (Uniform r.v.)

$f_X(x) = \dfrac{d F_X(x)}{dx}$



Area under the p.d.f. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

$F_X(-\infty) = 0$, $F_X(\infty) = 1$

$$F_X(x) = \int_{-\infty}^{x} f_X(\tilde{x})\,d\tilde{x} = \int_{-\infty}^{x} f_X(v)\,dv = \int_{-\infty}^{x} f_X(x)\,dx$$

Expected Value of a r.v.

Jan 22 3

$$E[X] \triangleq \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$

, Mean value

, Center of mass

Ex|| Cont'd $E[X] = \int_0^2 x \frac{1}{2}\, dx = \frac{x^2}{4}\Big|_0^2 = \frac{4}{4} - 0 = 1$

$P(X=1) = 0$ probability. , $f_X(1) = \frac{1}{2}$ (likelihood)

Ex|| Discrete r.v. $X(H) = 2$, $X(T) = 0$, $P(H) = P(T) = \frac{1}{2}$.

$$f_X(x) = \frac{1}{2}\delta(x) + \frac{1}{2}\delta(x-2)$$

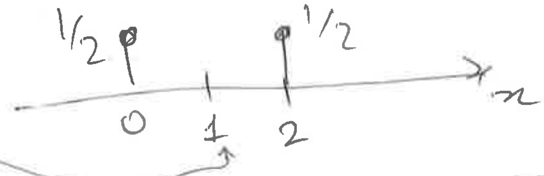
$$E[X] = \int x \frac{\delta(x)}{2}\, dx + \int \frac{1}{2} x\, \delta(x-2)\, dx$$

$$E[X] = 0 + \frac{2}{2} = 1$$

Discrete r.v.

$$E[X] = \sum_i x_i P_X(X=x_i) = \sum_{i=0}^{1} x_i P(X=x_i)$$

$$E[X] = 0 \cdot P(X=0) + 2\, P(X=2) = 2\,\frac{1}{2} = 1$$



Variance of a r.v.

$$\mathrm{Var}(X) = E[(X-\mu_X)^2] \quad \text{where } \mu_X = E[X]$$

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\, dx \quad \text{(Cont. r.v.)}$$

$$\mathrm{Var}(X) = \sum_i (x_i - \mu_X)^2 P(X=x_i) \quad \text{(Discrete r.v.)}$$

Jan 22/4

Ex|| Calculate the Variance of the $U[0,2]$ (Uniform r.v $[0,2]$)

$\mu_x = 1 = E[X]$

$$Var(X) = \int_0^2 (x-1)^2 \frac{1}{2}\, dx = \cdots$$

Ex|| Calculate the variance of the discrete r.v.

	x_i	$P(X)$
$i=0$	0	1/2
$i=1$	2	1/2

$$Var(X) = E\left[(X-\mu_x)^2\right] = \sum_{i=0}^{1} (x_i - 1)^2 P(X=x_i)$$

$\mu_x = E[X] = 1$

$$= 1^2 \tfrac{1}{2} + (2-1)^2 \tfrac{1}{2}$$

$$= \tfrac{1}{2} + \tfrac{1}{2} = 1$$

$\sigma^2 = Var(X)$.

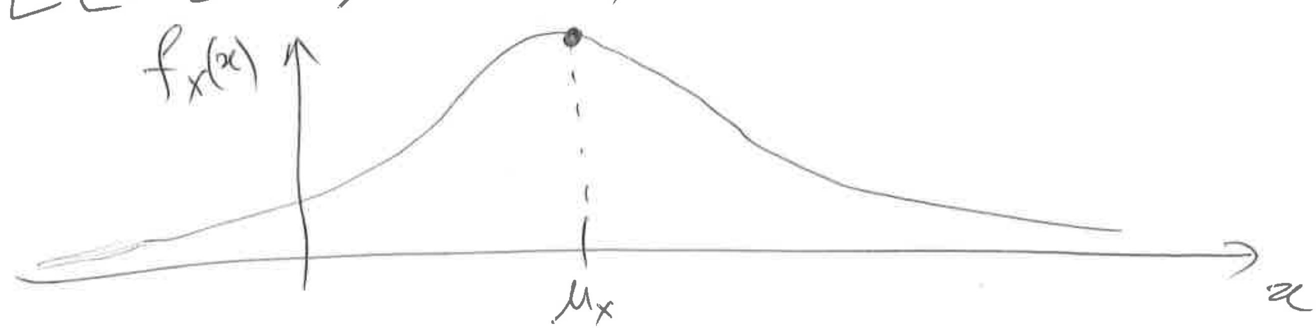
Standard Deviation : σ.

If $\sigma \downarrow$ the pdf is concentrated around the mean.

Gaussian p.d.f. (Normal r.v.)

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\frac{(x-\mu_x)^2}{\sigma^2}}$$

$E[X] = \mu_x$ / $Var(X) = \sigma_x^2$



Discrete r.v.'s. Bernoulli, Binomial.

Bernoulli : Cointoss r.v.

$$P(X = \overset{1}{\text{success}}) = p$$
$$P(X = \underset{0}{\text{fail}}) = 1-p = q.$$

$$p+q=1$$
$$\sum_i P(X=x_i) = 1$$

Binomial: Toss the coin several-n times:(n) & count the number of successes.

r.v. Y can take $Y=0,\ Y=1,\ \ldots,\ Y=n$

indepence

$$P(Y=0) = P(X=0)\,P(X=0)\cdots P(X=0)$$
$$= q \times q \times \cdots \times q = q^n$$

$$P(Y=k) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{(n-k)!\,k!}\, p^k q^{n-k}.$$