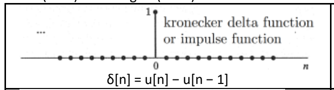
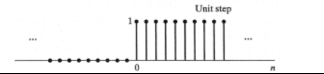
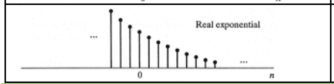
Classical Signal Chp 1
Impulse Functions

Note: (n-n0) = shift right. (n+n0) = shift left

 δ[n] = u[n] – u[n – 1]	$\delta[n] = \begin{cases}1, n=0\\0, n\neq 0\end{cases}$
	$u[n] = \begin{cases}1, n\geq 0\\0, n<0\end{cases}$
	$x[n] = a^n$, $0 < a < 1$

Discrete Convolution
- $y[n] = x[n] * h[n] = \sum_k x[n]h[n-k] = \sum_k x[n-k]h[k]$ → commutative

Discrete Time Systems
- Transformation T of input sequence $x[n]$ into output sequence $y[n]$
- $y[n] = T(x[n])$

Properties of Systems
Stable
- Bounded input, bounded output: $\sum_n |h[n]| < \infty$
- **FIR filters:** stable as long as values are finite → finite sum of numbers are finite
- $h[n]$ is bounded if $|h[n]| < B_h$ for all n. $y[n]$ is bounded if $|y[n]| < B_y$ for all n
- System is stable if $x[n]$ is bounded, then $y[n]$ is also bounded

$\boldsymbol{y[n] = g(x)x(n)}$ → if g(n) bounded	$\boldsymbol{y[n] = x[n-n_0]}$
$\boldsymbol{y[n] = x[n] + 3u[n+1]}$ → $\lvert x[n]\rvert \leq B_x$ and $y[n] \leq B_x + 3u[n+1] \leq B_x + 3$	$\boldsymbol{y[n] = e^{x[n]}}$ $\lvert x[n]\rvert \leq B_x$ and $\lvert y[n]\rvert \leq e^{B_x}$

Causal
- $y[n]$ does not depend on future inputs, i.e. for $n > n_0$ → $-\infty < n < \infty$
- Output $y[n]$ at time $n = n_0$ only depends on input $x[n]$ at $n \leq n_0$
- **LTI system only:** is causal only when $h[n] = 0$ for $n < 0$

$\boldsymbol{y[n] = g(x)x(n)}$	$\boldsymbol{y[n] = x[n-n_0]}$
$\boldsymbol{y[n] = e^{x[n]}}$	causal: $n - n_0 \leq n$, then $n_0 \geq 0$.
$\boldsymbol{y[n] = x[n] + 3u[n+1]}$	not causal: $n - n_0 > n$, then $n_0 < 0$.

Linear
- If $T(ax_1[n] + bx_2[n]) = ay_1[n] + by_2[n]$, it is linear
- Linear combination of inputs = linear combination of corresponding outputs

$\boldsymbol{y[n] = g(x)x(n)}$	$\boldsymbol{y[n] = e^{x[n]}}$
$\boldsymbol{y[n] = x[n-n_0]}$	$T(ax_1[n] + bx_2[n])$
$\boldsymbol{y[n] = x[n] + 3u[n+1]}$	$= e^{ax_1[n]+bx_2[n]} \neq ay_1[n] + by_2[n]$

Time invariant
- $x_1[n-n_0] * h[n] = \sum_{k=-\infty}^{\infty} x_1[k-n_0]h[n-k]$
  $= \sum_{m=-\infty}^{\infty} x_1[m]h[n-n_0-m]$
  $= y_1[n-n_0]$

$\boldsymbol{y[n] = g(x)x(n)}$ $T(x[n-n_0])$ $= g[n-n_0]x[x-n_0] = y[n-n_0]$	$\boldsymbol{y[n] = x[n-n_0]}$ $\boldsymbol{y[n] = e^{x[n]}}$
$\boldsymbol{y[n] = x[n] + 3u[n+1]}$ $(x[n-n_0]) = x[n-n_0] + 3u[n+1]) \neq y[n-n_0]$	

Memoryless
- $y[n]$ at any value of n depends on $x[n]$ only at time n.
- Output considers present input only
- A special case of causal (if memoryless, always causal)

$\boldsymbol{y[n] = g(x)x(n)}$	$\boldsymbol{y[n] = x[n-n_0]}$ **Only If $n_0 = 0$, Memoryless.**
$\boldsymbol{y[n] = e^{x[n]}}$	$\boldsymbol{y[n] = x[n] + 3u[n+1]}$ $y[n]$ at all values of n depends only on $x[n]$ at time n.

Properties of LTI systems
- All LTI systems can be expressed with convolution with impulse $h[n]$
- To find $h[n]$, set input $x[n] = \delta[n]$ and output $y[n] = h[n]$

Geometric Series: $a_n = ar^{n-1}$
- **Finite:** $a^0 + a + ar^2 + .. + ar^{n-1} = \frac{a(1-r^n)}{1-r}$ OR $\frac{a(r^n-1)}{r-1}$ where $r \neq 1$
- **Infinite:** $a^0 + a^1 + ar^2 + ar^3 + \cdots = \frac{a}{1-r}$ where $r < 1$
- **Infinite 2:** $\sum_{k=m}^{\infty} ar^k = \frac{ar^m}{1-r}$
- **Variation (finite):** $\sum_{k=N_1}^{N_2} a^k = \frac{a^{N_1} - a^{N_2+1}}{1-a}$ where $N_2 \geq N_1$
- **Variation (finite, complex):** $\sum_{k=N_1}^{N_2} e^{j\omega k} = \frac{e^{j\omega N_1} - e^{j\omega(N_2+1)}}{1-e^{j\omega}}$ where $N_2 \geq N_1$

Complex Numbers Identities

Cartesian: $z = a + jb$	**Polar:** $z = \lvert z\rvert e^{j\varphi}$	$\lvert z\rvert = \sqrt{a^2+b^2}$
Euler: $e^{jx} = \cos x + j\sin x$ → $\cos x = \frac{e^{jx}+e^{-jx}}{2}$ → $\sin x = \frac{e^{jx}-e^{-jx}}{2j}$		

Continuous / Discrete Time Fourier Transform
CFT: $\int_{-\infty}^{\infty} x(t)e^{-j\Omega t}dt$

Inverse CFT: $x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\Omega)e^{j\Omega t}d\Omega$

DTFT: $X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{j\omega n}$

Inverse DTFT: $x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega n}d\omega$

Classical Signal Chp 2
Fourier Conversion

Sequence	Fourier Transform
$\boldsymbol{\delta(n)}$	$\boldsymbol{1}$
$\boldsymbol{\delta(n-n_0)}$ where $n_0 \in \mathbb{Z}$	$\boldsymbol{e^{-j\omega n_0}}$ → $\boldsymbol{n_0}$ **is an integer**
$\boldsymbol{1}$ where $(-\infty < n < \infty)$	$\sum_{k=-\infty}^{\infty} 2\pi\delta(\omega + 2\pi k)$
$\boldsymbol{a^n u[n]}$ where $(\lvert a\rvert < 1)$	$\frac{1}{1-ae^{-j\omega}}$
$\boldsymbol{u[n]}$	$\frac{1}{1-e^{-j\omega}} + \sum_{k=-\infty}^{\infty} \pi\delta(\omega + 2\pi k)$
$\boldsymbol{(n+1)a^n u[n]}$ where $(\lvert a\rvert < 1)$	$\frac{1}{(1-ae^{-j\omega})^2}$
$\frac{r^n \sin\omega_p(n+1)}{\sin\omega_p}\boldsymbol{u[n]}$ where $(\lvert r\rvert < 1)$	$\frac{1}{1-2r\cos\omega_p e^{-j\omega} + r^2e^{-j2\omega}}$
$\frac{\sin\omega_c n}{\pi n}$	$X(e^{j\omega}) = \begin{cases}1, & \lvert\omega\rvert < \omega_c\\ 0, & \omega_c < \lvert\omega\rvert \leq \pi\end{cases}$
$x[n] = \begin{cases}1, & 0 \leq n \leq M\\ 0, & otherwise\end{cases}$	$\frac{\sin\left[\frac{\omega(M+1)}{2}\right]}{\sin\left(\frac{\omega}{2}\right)}e^{-\frac{j\omega M}{2}}$
$\boldsymbol{e^{j\omega_0 n}}$	$\sum_{k=-\infty}^{\omega} 2\pi\delta(\omega - \omega_0 + 2\pi k)$
$\boldsymbol{\cos(\omega_0 n + \phi)}$	$\sum_{k=-\infty}^{\infty}[\pi e^{j\phi}\delta(\omega - \omega_0 + 2\pi k) + \pi e^{-j\phi}\delta(\omega + \omega_0 + 2\pi k)]$

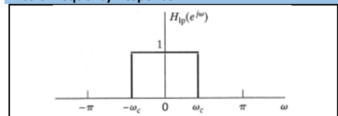
Fourier Equations (Simplification)

$$\frac{1}{(1-ae^{-j\omega})(1-be^{-j\omega})} = \frac{a/(a-b)}{1-ae^{-j\omega}} - \frac{b/(a-b)}{1-be^{-j\omega}}$$

Properties of DTFT

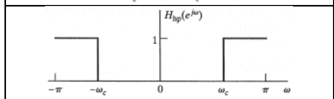
Linearity	$ax_1(n) + bx_2(n) \leftrightharpoons aX_1(e^{j\omega}) + bX_2(e^{j\omega})$
Time shift	$x(n-n_0) \leftrightharpoons e^{-j\omega n_0}X(e^{j\omega})$
Frequency Shift	$e^{j\omega n_0}x(n) \leftrightharpoons X(e^{j(\omega-w_0)})$
Convolution Theorem 1	$x(n) * h(n) \leftrightharpoons X(e^{j\omega})H(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega})$
Theorem 2	$x(n)h(n) \leftrightharpoons X(e^{j\omega}) * H(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\theta})H(e^{j(\omega-\theta)})d\theta$ (if periodic, else is infinity)

Filters Frequency Response



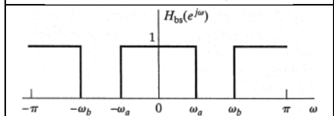
Ideal Low-Pass Filter

$H_{lp}(e^{j\omega}) = \begin{cases}1, & \lvert\omega\rvert < \omega_c\\ 0, & \omega_c < \lvert\omega\rvert \leq \pi\end{cases}$



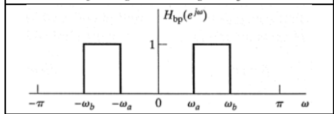
Ideal High-Pass Filter

$H_{hp}(e^{j\omega}) = \begin{cases}0, & \lvert\omega\rvert < \omega_c\\ 1, & \omega_c < \lvert\omega\rvert \leq \pi\end{cases}$



Ideal Band-stop Filter

$H_{bs}(e^{j\omega}) = \begin{cases}1, & \lvert\omega\rvert < \omega_a\\ 0, & \omega_a < \lvert\omega\rvert < \omega_b\\ 1, & \omega_b < \lvert\omega\rvert < \pi\end{cases}$



Ideal Bandpass Filter

$H_{bp}(e^{j\omega}) = \begin{cases}0, & \lvert\omega\rvert < \omega_a\\ 1, & \omega_a < \lvert\omega\rvert < \omega_b\\ 0, & \omega_b < \lvert\omega\rvert < \pi\end{cases}$

**Find ω to prove filter

Convert HP <-> LP : $H_{hp} = 1 - H_{lp}$

Convert BP <-> BS: $H_{bp} = 1 + H_{bs}$ → made up of 2 LP

Convert BS <-> HP: $H_{bs} = H_{hp} * H_{bsf}$

Filter Design by Windowing
– **Start with ideal frequency response $H(e^{j\omega}) \leftrightarrow h_I[n]$**

– **Multiply by window w[n] in time domain: $h_I[n]w[n]$**

– **Window shape and length M tradeoffs ringing and cutoff sharpness**

Classical Signal Chp 3
Phase Distortion
Consider ideal delay system: $h[n] = \delta[n = n_d]$
- $H(e^{j\omega}) = e^{-j\omega n_d} \to \lvert H(e^{j\omega})\rvert = 1 \to angle = -\omega n_d$

Ideal filter with linear phase acceptable as filter without phase distortion
- **E.g. ideal linear phase LPF:** $H_{lp} = \begin{cases}e^{-j\omega\alpha}, & \lvert\omega\rvert < \omega_c, \alpha\epsilon R\\ 0, & \omega_c < \lvert\omega\rvert \leq \pi\end{cases}$
- **Generalised linear phase:** $\omega\alpha + \beta$

Type 1/2 FIR Linear Phase System
Symmetric causal FIR $h[n] = \begin{cases}0, & n < 0, n > M\\ h[M-n], & 0 \leq n \leq M\end{cases}$

even M: $\boldsymbol{H(e^{j\omega}) = \sum_{n=0}^{M} h[n]e^{-j\omega n} = e^{-j\omega M/2}\sum_{k=0}^{M/2} a[k]\cos\omega k}$

Where $a[0] = h[M/2]$ and $a[k] = 2h[(M/2) - k]$

Delay $\alpha = M/2$ (an integer)

odd M: $\boldsymbol{H(e^{j\omega})\sum_{n=0}^{M} h[n]e^{-j\omega n} = e^{-j\omega M/2}\sum_{k=0}^{(M+1)/2} b[k]\cos\left[\omega\left(k - \frac{1}{2}\right)\right]}$

where $b[k] = 2h[(M+1)/2 - k)]$

Delay $\alpha = M/2$ (an integer + half)

Not Suitable for HPF and BSF because $H(e^{j\omega}) = 0$

Kaiser Window

$$w[n] = \begin{cases}\frac{I_o\left[\beta\left(1 - \left[\frac{n-M/2}{M/2}\right]^2\right)^{\frac{1}{2}}\right]}{I_o(\beta)}, & 0 \leq n \leq M\\ 0, & otherwise\end{cases}$$

Picking M and β for Kaiser Window

$\beta = \begin{cases}0.1102(A-8.7), & A > 50\\ 0.5842(A-21)^{0.4} + 0.07886(A-21), & 21 \leq A \leq 50\\ 0.0, & A < 21\end{cases}$

β controls shape of window:

If **$\beta = 0$,** Kaiser window is rectangular window

If **β increases,** Kaiser window looks more like Gaussian

$M = \frac{A-8}{2.285\Delta\omega}$	$A = -20\log_{10}\delta$	**For LP:** $\Delta\omega = \omega_s - \omega_p$ **For HP:** $\Delta\omega = \omega_p - \omega_s$
$\omega_c = \frac{\omega_p + \omega_s}{2}$		
Finding magnitude: $\boldsymbol{E.g.\ 0.95 < \lvert H(e^{j\omega})\rvert < 1.05}$ $Magnitude = \frac{Max + Min}{2} = \frac{0.95 + 1.05}{2}$		

We wish to use Kaiser window method to design a FIR filter with generalized linear phase that meets the following specifications:

Mag = 1 $\boldsymbol{0.95 < \lvert H(e^{j\omega})\rvert < 1.05}$ H	$\boldsymbol{0 \leq \lvert\omega\rvert \leq 0.2\pi}$
Mag = 0 $\boldsymbol{\lvert H(e^{j\omega})\rvert < 0.1}$ L	$\boldsymbol{0.3\pi \leq \lvert\omega\rvert \leq 0.6\pi}$
Mag = 1 $\boldsymbol{0.95 < \lvert H(e^{j\omega})\rvert < 1.05}$ H	$\boldsymbol{0.7\pi \leq \lvert\omega\rvert \leq \pi}$

Find the Kaiser window parameters β and M and so that the resulting windowed filter satisfies the above criteria. Justify your choice of δ and Δw. What is the frequency response of the ideal filter?

$\delta = 0.05$ → because $\delta = 0.05$ is stricter than $\delta = 0.1$ in stopband

$A = -20\log_{10}0.05 = 26.0206$

$\beta = 0.5842(A-21)^{0.4} + 0.07886(A-21) = 1.5099$

$\Delta\omega_1 = 0.3\pi - 0.2\pi = 0.1\pi$ and $\Delta\omega_2 = 0.7\pi - 0.6\pi = 0.1\pi$ → same value

$M = \frac{A-8}{2.285\Delta\omega} = 25.1$ → round up $M = 26$

$$H_I(e^{j\omega}) = \begin{cases}e^{-j\omega M/2}, & 0 \leq \lvert\omega\rvert \leq \frac{0.2+0.3}{2}\pi\\ 0, & \frac{0.2+0.2}{2}\pi \leq \lvert\omega\rvert \leq \frac{0.6+0.7}{2}\pi\\ 2e^{j\omega M/2}, & \frac{0.6+0.7}{2}\pi \leq \lvert\omega\rvert \leq \pi\end{cases}$$

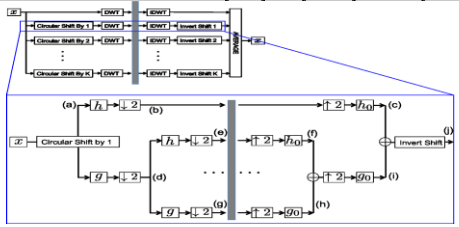
$$= \begin{cases}e^{-j\omega M/2}, & 0 \leq \lvert\omega\rvert \leq 0.25\pi\\ 0, & 0.25\pi \leq \lvert\omega\rvert \leq 0.65\pi\\ 2e^{-j\omega M/2}, & 0.65\pi \leq \lvert\omega\rvert \leq \pi\end{cases}$$

Discrete Fourier Transform (DFT)
– Sampled version of DTFT with sampling period $2\pi/N$

– Can be computed in N log N operations with FFT

– Efficiently compute DTFT and convolutions

Classical Signal Chp 4
Wavelet Consider one cycle of the Spincycle transform illustrated. Suppose $x = [6\ \ 12\ \ 120\ \ 116]$. What are the coefficients at location **(a)-(j)** in the spincycle, assuming we use the Haar transform, where $g = g_0 = \left[\frac{\sqrt2}{2}\ \frac{\sqrt2}{2}\right]$, $h = \left[-\frac{\sqrt2}{2}\ \frac{\sqrt2}{2}\right]$ and $h_0 = \left[\frac{\sqrt2}{2}\ -\frac{\sqrt2}{2}\right]$



(a) $Required\ coefficient\ 1 = [116\ \ 6\ \ 12\ \ 120]$ → shift right

(b) $Required\ coefficient\ 2 = [116\ \ 6\ \ 12\ \ 120] * \left[-\frac{\sqrt2}{2}\ \frac{\sqrt2}{2}\right]$ → ↓ 2
$= \sqrt2\left[\frac{116-6}{2}\ \ \frac{12-120}{2}\right]$
$= \sqrt2[55\ \ -54]$

(c) $Required\ coefficient\ 3i = \sqrt2[55\ \ 0\ \ -54\ \ 0]$
$Final\ Required\ coefficient\ 3 = \sqrt2[55\ \ 0\ \ -54\ \ 0] * \left[\frac{\sqrt2}{2}\ -\frac{\sqrt2}{2}\right]$
$= \sqrt2\sqrt2[27.5\ \ -27.5\ \ -27\ \ 27]$
$= [55\ \ -55\ \ -54\ \ 54]$

(d) $Required\ coefficient\ 4 = [116\ \ 6\ \ 12\ \ 120] * \left[\frac{\sqrt2}{2}\ \frac{\sqrt2}{2}\right]$ → ↓ 2
$= \sqrt2\left[\frac{116+6}{2}\ \ \frac{12+120}{2}\right]$
$= \sqrt2[61\ \ 66]$

(e) $Required\ coefficient\ 5 = \sqrt2[61\ \ 66] * \left[-\frac{\sqrt2}{2}\ \frac{\sqrt2}{2}\right]$ → ↓ 2
$= \sqrt2\sqrt2\left[\frac{61-66}{2}\right]$
$= [-5]$

(f) $Required\ coefficient\ 6 = [-5\ \ 0] * \left[\frac{\sqrt2}{2}\ -\frac{\sqrt2}{2}\right]$
$= \sqrt2[-2.5\ \ 2.5]$

(g) $Required\ coefficient\ 7 = \sqrt2[61\ \ 66] * \left[\frac{\sqrt2}{2}\ \frac{\sqrt2}{2}\right]$ → ↓ 2
$= \sqrt2\sqrt2\left[\frac{61+66}{2}\right]$
$= [127]$

(h) $Required\ coefficient\ 8 = [127\ \ 0] * \left[\frac{\sqrt2}{2}\ \frac{\sqrt2}{2}\right]$
$= \sqrt2[63.5\ \ 63.5]$

(i) $Required\ coefficient\ 9i = \sqrt2[-2.5\ \ 2.5] + \sqrt2[63.5\ \ 63.5]$
$= \sqrt2[61\ \ 66]$
$Final\ Required\ coefficient\ 9 = \sqrt2[61\ \ 0\ \ 66\ \ 0] * \left[\frac{\sqrt2}{2}\ \frac{\sqrt2}{2}\right]$
$= \sqrt2\sqrt2[30.5\ \ 30.5\ \ 33\ \ 33]$
$= [61\ \ 61\ \ 66\ \ 66]$

(j) $Required\ coefficient\ 10i = [55\ \ -55\ \ -54\ \ 54] + [61\ \ 61\ \ 66\ \ 66]$
$= [116\ \ 6\ \ 12\ \ 120]$
$Final\ Required\ coefficient\ 10 = [6\ \ 12\ \ 120\ \ 116]$

Statistical Chp 3 - Random walk
$\boldsymbol{x_n: x_N = x_0 + \sum_{n=1}^{N} z_n}$

$\boldsymbol{x_0}$: initial position

$\boldsymbol{x_n} = x_{n-1} + z_n$ for $n \geq 1$

$\boldsymbol{z_n} = \pm 1$ with equal probability & Independent and Identically Distributed (i.i.d.)

$\boldsymbol{p(x_n = k) = p\left(a = \frac{n+k}{2}\ heads\ in\ independent\ coin\ tosses\right)}$

$\boldsymbol{= \binom{n}{(n+k)/2}\frac{1}{2n}}$

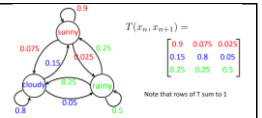
Markov Process: Random process is Markov if future and past conditionally independent given present

Markov Chain:

$p(x_1, \ldots, x_{n-1}, x_{n+1}, \ldots, x_N \mid x_N)$
$= p(x_1, \ldots, x_{n-1} \mid x_n)p(x_{n+1}, \ldots x_N \mid x_n)$

$\boldsymbol{x_n}$: state at timept n

Transition Matrix: rows of T sum to 1



3 state Markov chain example

Markov Chain Properties
Property (1): $\boxed{\boldsymbol{\pi_{n+1} = \pi_n T}}$ where π_n is a row vector & $\pi_n(i) = p(x_n = s_i)$
$\pi_{n+1}(j) = p(x_{n+1} = s_j)$

Property (2): Stationary Distribution

$\pi_{n+1} = \pi_n T$ → $\pi_{n+2} = \pi_{n+1}T = \pi_n T^2$ → …→ $\boxed{\boldsymbol{\pi_{n+k} = \pi_n T^k}}$

Eigenvector ($\boldsymbol{\pi^*}$): If $\pi^* = \pi^* T$, then π^* is stationary distri. of T where $\pi^* = 1$.

Once we reach π^*, distribution of the state stays constant

Fundamental Theorem of Markov Chain Satisfy: If all entries in $T^n > 0$ for all $n \geq 1$, fundamental theorem applies & T has a unique stationary distribution.

E.g. (does not satisfy): (1) identity matrix: no unique stationary distribution.

(2) Random walk: x_n does not satisfy the Fundamental Theorem. →

$p(x_{n0} = n_0 + 1) = 0$ (e.g., at time 5, there is zero probability of $x_5 = 6$).

Statistical Chp 1 – Random Variable

Cont. x	pdf	$0 \le p(x) \le \infty$	$\int_{-\infty}^{\infty} p(x) = 1$
Discrete x	pmf	$0 \le p(x) \le 1$	$\sum_x p(x) = 1$

Joint Distributions $p(X,Y)$

If x & y *continuous*: joint probability distribution of x & y
If x & y *discrete*: joint probability mass function of x & y
Valid $p(x,y)$: $\int_y \int_x p(x,y)\,dx dy = 1$, (2) or $p(x,y) \ge 0$ for all x, y

Marginalisation

Purpose: recover probability distribution of any variable in joint distribution

Cont. x, y	$p(x) = \int_y p(x,y)\,dy$	$p(y) = \int_x p(x,y)\,dx$

Conditional Probability

$p(x\|y) = \frac{p(x,y)}{p(y)}$	$p(x\|y = y^*) = \frac{p(x,y=y^*)}{p(y=y^*)} = \frac{p(x,y=y^*)}{\int P(x,y=y^*)dx}$	
$p(x,y) = p(y)p(x\|y)$	$p(y,x) = p(x)p(y\|x)$	

Bayes' Rule

P(x|y)	$p(x\|y) = \frac{p(x)p(y\|x)}{\int_x p(x)p(y\|x)dx} = \frac{p(x,y)}{p(y)}$
P(y|x)	$p(y\|x) = \frac{p(y)p(x\|y)}{\int_y p(y)p(x\|y)dx} = \frac{p(y,x)}{p(x)}$

Independence – if x & y are independent.

$p(x\|y) = p(x)$ or $p(y\|x) = p(y)$	$p(x,y) = p(x)p(y\|x) = p(x)p(y)$

Expectation – tells us about average

Single Variable: mean, variance.

E	Cont. x	$E[f(x)] = \int_x f(x)p(x)dx$
C	Cont. x	$E[(x-\mu_x)^2] = \int_x (x-\mu_x)^2 p(x)$

Two Variables

E	Cont. x	$E[f(x,y)] = \int_x \int_y f(x,y)p(x,y)\,dxdy$
C	Cont. x	$E[(x-\mu_x)(y-\mu_y)] = \int_x \int_y (x-\mu_x)(y-\mu_y)\,dxdy$

Covariance: measures presence of linear relationship
$Cov(x,y) = 0$: does not imply x & y are independent.
$Cov(x,y) \ne 0$: linear relationship → x & y not independent

Rules of Expectation

$E[af(x) + bg(x)] = aE(f(x)) + bE(g(x))$ → Linearity of expectation
$Cov(x,y) = E(x,y) - E(x)E(y)$ → can be -ve
Hence, $Var(x) \equiv Cov(x,x) = E(x^2) - (E(x))^2$
If x & y independent, $E[f(x)g(y)] = E[f(x)]E[g(y)]$ → $Cov(x,y) = 0$

Conditional Expectation

Cont. x	$E_{p(x\|y)}[f(x,y)] \triangleq \int_x f(x,y)p(x\|y)dx$

Tut 1 qn 1 Consider random variables x and y with joint distribution,
$$p(x,y) = \begin{cases} x+y, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & otherwise \end{cases}$$
(a) Verify that $p(x,y)$ is a valid probability distribution
(b)Find $p(x)$, $p(y)$ and $p(x|y)$ (e) Find $Var(x)$ and $Var(y)$
(c)Are x and y independent? Explain (f) Find $Cov(x,y)$
(d)Find $E(x)$ and $E(y)$ (g) Find $E(x^2|y = 0.5)$

(a) to verify, $\int_y \int_x p(x,y)\,dxdy = 1$
$\int_0^1 \int_0^1 (x+y)\,dxdy = \int_0^1 x\,dx + \int_0^1 y\,dx\,dy = \cdots = \frac{1}{2} + \frac{1}{2} = 1$ **(proven)**

(b) $p(x) = \int_y p(x,y)\,dy = \int_0^1 (x+y)dy = \left[x + \frac{y^2}{2}\right]_0^1 = x + \frac{1}{2}$
$p(y) = \int_x p(x,y)\,dx = \int_0^1 (x+y)dx = \left[\frac{x^2}{2} + y\right]_0^1 = \frac{1}{2} + y$
$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{x+y}{0.5+y}$ for $0 \le (x,y) \le 1$

(c) x and y are not independent → $p(x|y) \ne p(x)$

(d) $E(x) = \int_x (x)p(x)\,dx = \int_0^1 (x)(x + \frac{1}{2})dx = \cdots = \frac{7}{12}$
$E(y) = \int_y (y)p(y) = \int_0^1 (y)\left(\frac{1}{2} + y\right)dy = \cdots = \frac{7}{12}$

(e) $Var(x) = E(x^2) - (E(x))^2 = \int_0^1 (x^2)p(x)dx - \left(\frac{7}{12}\right)^2 = \cdots = \frac{11}{144}$
$Var(y) = E(y^2) - (E(y))^2 = \int_0^1 (y^2)p(y)dy - \left(\frac{7}{12}\right)^2 = \cdots = \frac{11}{144}$

(f) $Cov(x,y) = E(x,y) - E(x)E(y)$
$E(x,y) = \int_x \int_y (xy)(x+y)\,dxdy = \int_0^1 \int_0^1 (x^2y + xy^2)dxdy = \cdots = \frac{1}{3}$
$Cov(x,y) = \frac{1}{3} - \left(\frac{7}{12}\right)\left(\frac{7}{12}\right) = -\frac{1}{144}$

(g) $E(x^2|y = 0.5) = \int_x (x^2)p(x|y = 0.5)\,dx = \int_0^1 x^2 \left(\frac{x+0.5}{0.5+0.5}\right)dx$
$= \int_0^1 x^2(x + 0.5)dx = \frac{5}{12}$

Tut 1 qn 5: if two random variables x & y are independent, then $Cov(x,y) = 0$. Give an example where $Cov(x,y) = 0$, but x and y are not independent.
$p(5,5) = \frac{2}{5}$ $p(-5,-5) = \frac{2}{5}$, and $p(-10,10) = \frac{1}{10}$ $p(10,-10) = \frac{1}{10}$
$p(x,y) = 0$ otherwise → x and y not independent (knowing one determine the other)
$E(X,Y) = \sum_x \sum_y xyp(x,y) = (25)\left(\frac{2}{5}\right) + (25)\left(\frac{2}{5}\right) + (100)\left(\frac{1}{10}\right) + (100)\left(\frac{1}{10}\right) = 0$
$E(x) = \sum_x xp(x) = (5)\left(\frac{2}{5}\right) + (-5)\left(\frac{2}{5}\right) + (-10)\left(\frac{1}{10}\right) + (10)\left(\frac{1}{10}\right)$
$E(y) = \sum_y yp(y) = (5)\left(\frac{2}{5}\right) + (-5)\left(\frac{2}{5}\right) + (-10)\left(\frac{1}{10}\right) + (10)\left(\frac{1}{10}\right)$
$Cov(x,y) = E(XY) - E(x)E(y) = 0$ → $Cov(x,y) = 0$ despite being dependent

Statistical Chp 2 – Probabilistic Signal Detection

Conditional Probability (variables)

2	$p(x,y) = p(x)p(y\|x)$
3	$p(x_1,x_2,x_3) = p(x_1)p(x_2,x_3\|x_1) = p(x_1)p(x_2\|x_1)p(x_3\|x_1,x_2)$
N	$p(x_1,\ldots,x_N) = p(x_1)p(x_2,\ldots,x_N\|x_1) = p(x_1)p(x_2\|x_1)p(x_3,\ldots,x_N\|x_1,x_2) = p(x_1)p(x_2\|x_1)p(x_3\|x_1,x_2)\ldots p(x_N\|x_1,\ldots,x_{N-1})$

Estimating x given observation y

Max likelihood: $\boxed{x_{ML} = argmax_x\, p(x = x^*|y)}$ → $p(observation|estimate)$
Maximum-A-Posteriori: $\boxed{x_{MAP} = argmax_x p(y|x = x^*)}$ → $p(estimate|observation)$
Min mean square error estimate: $\boxed{x_{MMSE} = argmin_{\hat{x}} E_{p(x|y)}(x - \hat{x})^2 = E_{p(x|y)}(x)}$
Note: $\boxed{argmax\,(af(x)) = argmax(f(x))}$
Note2: MAP minimize P of error, MMSE minimize average

Tut 2 qn 1: Suppose random variables x1 and x3 are conditionally independent given x2. Show that p(x1, x2, x3) = p(x1)p(x2|x1)p(x3|x2) = p(x3)p(x2|x3)p(x1|x2)

$p(x_1,x_2,x_3)$	$p(x_3,x_2,x_1)$
$= p(x_1)p(x_2,x_3\|x_1)$	$= p(x_3)p(x_2,x_1\|x_3)$
$= p(x_1)p(x_2\|x_1)p(x_3\|x_1,x_2)$	$= p(x_3)p(x_2\|x_3)p(x_1\|x_3,x_2)$
$= p(x_1)p(x_2\|x_1)p(x_3\|x_2)$	$= p(x_3)p(x_2\|x_3)p(x_1\|x_2)$
$p(x_1,x_2,x_3) = p(x_1)p(x_2\|x_1)p(x_3\|x_2) = p(x_3)p(x_2\|x_3)p(x_1\|x_2)$ (shown)	

Tut 2 qn 2: A coin has a random bias q with pdf p(q) = 2q for 0 ≤ q ≤ 1 (0 otherwise), where q is the probability of the coin turning up head. q is random and unknown, but assumed to be fixed for the duration of the following random experiment: the coin is flipped 10 times independently, resulting in 9 heads → observation. Compute ML estimate of q, MAP estimate of q, MMSE estimate of q.

$q_{ML} = argmax_q p(x = 9|q)$
$= argmax_q \binom{10}{9}(q)^9(1-q)^1$
$= argmax(q^9 - q^{10})$
$\frac{\partial}{\partial q}(q^9 - q^{10}) = 9q^8 - 10q^9$
$9q^8 - 10q^9 = 0$
$\boxed{q_{ML} = \frac{9}{10}}$

$q_{MAP} = argmax_q p(q|x = 9)$
$p(q|x = 9)$
$= \frac{p(q,x)}{p(x)}$
$= \frac{p(x=9|q)p(q)}{p(x=9)}$ → conditional probability
$= \frac{\binom{10}{9}q^9(1-q)^1(2q)}{\int_0^1 p(q)p(x=9|q)dq}$ → $p(a) = \int p(a,b)$
$= \frac{20(q^{10} - q^{11})}{\int_0^1 q\binom{10}{9}(q)^9(1-q)dq}$
$= \cdots$
$= 132(q^{10} - q^{11})$
$\frac{\partial}{\partial q}(q^{10} - q^{11}) = 10q^9 - 11q^{10}$
$10q^9 - 11q^{10} = 0$ → $\boxed{q_{MAP} = \frac{10}{11}}$

$q_{MAP} = argmax_q p(q|x = 9)$
$= argmax_q \frac{p(q)p(x=9|q)}{p(x=9)} = argmax_q\, p(q)p(x = 9|q)$ →same argmax
$= argmax_q (2q)\binom{10}{9}q^9(1-q)$
$= argmax(q^{10} - q^{11})$ → $\frac{\partial}{\partial q}(q^{10} - q^{11}) = 0 \ldots.$ → $\boxed{q_{MAP} = \frac{10}{11}}$

$q_{MMSE} = E_{p(q|x=9)}(q) = \int_0^1 (q)132(q^{10} - q^{11})\,dq = \cdots = 11/13$

Tut 2 qn 3 Consider joint distribution, $p(x,y) = \begin{cases} x+y, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & otherwise \end{cases}$, with random x & y variables. Suppose I play the following game with a student from day n = 0 to $n = \infty$. On day n, I sample $(x_n, y_n) \sim p(x,y)$. Given y_n, the student has to guess (estimate) the value of x_n. Let the student's guess be $\hat{x}_n$. The student has to pay me $(x_n - \hat{x}_n)^2$ dollars. **(a)** To minimize the penalty (on average), should the student use ML, MAP or MMSE estimate? **(b)** Using the best strategy from (a), how much does the student pay me on average each day? **(c)** instead of the student paying me $(x_n - \hat{x}_n)^2$ dollars, I pay the student $\delta(x_n - \hat{x}_n)$ dollars. What is the optimal strategy the student should use to maximize the reward on average? How much would I pay the student on average each day? **(a)** use MMSE.

(b) $x_{n_{MMSE}} = E_{p(x_n|y_n)}(x)$
$= \int_0^1 (x_n)p(x_n|y_n)\,dx$
$= \int_0^1 x_n\left(\frac{x_n + y_n}{0.5 + y_n}\right)dx$ → derived from tut 1 qn 1b
$= \cdots$
$= \frac{1}{0.5+y}\left[\frac{1}{3} + \frac{1}{2}y\right]$ →subscript n was dropped to reduce clutter

Average payout $= \int_x \int_y p(x,y)[x_{MMSE} - x]^2\,dxdy$
$= \int_y p(y) \int_y p(x|y)(x_{MMSE} - x)^2\,dxdy$ → conditional probability
$= \int_y p(y)Var(x|y)dy$ → $x_{MMSE} = E(x|y)$

$Var(x|y) = E(x^2|y) - E(x|y)^2$
$E(x^2|y) = \int_x x^2 p(x|y)dx = \int_0^1 x^2\left(\frac{x+y}{0.5+y}\right)dx = \cdots = \frac{1}{0.5+y}\left[\frac{1}{4} + \frac{1}{3}y\right]$
$Var(x|y) = \frac{1}{0.5+y}\left[\frac{1}{4} + \frac{1}{3}y\right] - \left(\frac{1}{0.5+y_n}\left[\frac{1}{3} + \frac{1}{2}y_n\right]\right)^2$
$\int_y p(y)Var(x|y)dy = \int_0^1 (0.5 + y)\left(\frac{1}{0.5+y}\left[\frac{1}{4} + \frac{1}{3}y\right] - \left(\frac{1}{0.5+y_n}\left[\frac{1}{3} + \frac{1}{2}y_n\right]\right)^2\right)dy$
$= \cdots$
$= \mathbf{0.0757}$ (Student pays 0.0757 dollars on average each day)

(c) This is actually a trick question. The optimal strategy is technically the MAP strategy. However, because $p(x|y)$ is a continuous distribution and so no matter what the student guesses, he or she has 0 probability of guessing it correctly. Therefore, I pay the student 0 dollars on average each day despite the student using an optimal MAP strategy.

Pattern recognition Chp 1

Tut 1 qn 1: Find eigenvalues and eigenvectors for matrix $A = \begin{pmatrix} 5 & 3 \\ 1 & 3 \end{pmatrix}$.

$A - \lambda I = \begin{pmatrix} 5-\lambda & 3 \\ 1 & 3-\lambda \end{pmatrix}$ → $A - \lambda I = \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix}$
$\det(A - \lambda I) = 0$ → $(a-\lambda)(d-\lambda) - (bc) = 0$
$(5-\lambda)(3-\lambda) - (3)(1) = 0$ → $(a-\lambda)(d-\lambda) - (bc) = 0$
$\lambda^2 - 8\lambda + 12 = 0$ → $\lambda = 4 \pm 2$ → Matrix A has 2 $\boxed{\text{eigenvalues 2 and 6}}$
When $\lambda = 2$, $\begin{pmatrix} 5-2 & 3 \\ 1 & 3-2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ → $\begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
$\begin{pmatrix} 3x_1 + 3x_2 \\ x_1 + x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ → find a vector satisfying eq
$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ → normalize $\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$ → normalise to get unit eigenvector
When $\lambda = 6$, $\begin{pmatrix} 5-6 & 3 \\ 1 & 3-6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ → $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ → normalize $\boxed{\begin{pmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix}}$

Tut 1 qn 2: For $A = \begin{pmatrix} 2 & a \\ a & 2 \end{pmatrix}$, calculate eigenvalues of A and specify the values of A for which A is a valid covariance matrix.

$\det\begin{pmatrix} 2-\lambda & a \\ a & 2 \end{pmatrix} = 0$ → $(2-\lambda)(2-\lambda) - a^2 = 0$ → $\lambda = 2 \pm a$
The necessary and sufficient conditions for a covariance matrix is **symmetry and positive semidefinite** ($x^T Ax \ge 0$ for all x → $\lambda_m \ge 0$ for all m). Therefore, $-2 \le a \le 2$, so that eigenvalues are guaranteed to be at least 0 (+ve semidefinite).

Tut 1 qn 3: Consider the following set of data: $\begin{pmatrix} -2 \\ 3 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} -1 \\ 0 \end{pmatrix}$. What is the first principal axis? What is the first principal component of datapoint $x = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$? What is the reconstruction error from representing $x = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ using just first principal component?

$\mu = \frac{1}{4}\begin{pmatrix} -2+1+2-1 \\ 3+2-1+0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ → mean vector: $\mu = \frac{1}{N}\sum_n x^{(n)}$
$X = \left[\begin{pmatrix} -2 \\ 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ -2 \end{pmatrix}\begin{pmatrix} -1 \\ -1 \end{pmatrix}\right]$ → subtract mean vector from data points
$C = \frac{1}{4}\left[\begin{pmatrix} -2 \\ 2 \end{pmatrix}(-2 \quad 2) + \cdots + \begin{pmatrix} -1 \\ -1 \end{pmatrix}(-1 \quad -1)\right]$ → $C = \frac{1}{N}XX^T$
$= \begin{pmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{pmatrix}$
$\det\begin{pmatrix} 2.5-\lambda & -1.5 \\ -1.5 & 2.5-\lambda \end{pmatrix} = 0$ → $\lambda = 4$ or 1 → choose largest eigenvector
$(C - \lambda I) = \begin{pmatrix} -1.5 & -1.5 \\ -1.5 & -1.5 \end{pmatrix}$ → $\lambda = 4$
$\begin{pmatrix} -1.5 & -1.5 \\ -1.5 & -1.5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$ → $\begin{pmatrix} -1.5x_1 - 1.5x_2 \\ -1.5x_1 - 1.5x_2 \end{pmatrix} = 0$ → $(C - \lambda I)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$
$x_1 = -x_2$ → largest unit length principal axis: $v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$
1st principal component of $x = [1/\sqrt{2} \quad -1/\sqrt{2}]\left[\begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right] = \sqrt{2}$ → $v^T(x - \mu)$
Representation error: → $\left\|x(\mu + \sqrt{2}v_1)\right\|^2$
$\left\|\begin{pmatrix} 2 \\ 1 \end{pmatrix} - \left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \sqrt{2}\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}\right]\right\|^2 = \left\|\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\|^2 = 1^2 + 1^2 = 2$

Pattern recognition Chp 2

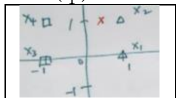
Tut 2 Qn 1: Given data samples x_1, x_2, x_3, x_4 to be 2,5,7,8. For Parzen's window $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. **(a)** what is the estimate of $p_h(x)$ at $x = 4$ for $h = 1, 2$ or 3? **(b)** what is the estimate of $p_h(x)$ at $x = 6$ for $h = 1, 2$ or 3? **(c)** Given data points x = 4 and 6 as the validation set, what is the best h (out of the possible values of 1, 2 and 3).

(a) Parzen's window estimation: $\boxed{p_h(x) = \frac{1}{N}\sum_{n=1}^N \frac{1}{h^d}\phi\left(\frac{x - x_n}{h}\right)}$
In this 1-D case, $d = 1$, $N = 4$ (4 data samples), and plugging $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$
New $\boxed{p_h(x) = \frac{1}{4}\sum_{n=1}^4 \frac{1}{h}\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(\frac{x-x_n}{h}\right)^2}{2}}}$
When $x = 4$, $h = 1$, $p_1(x = 4) = \frac{1}{4\sqrt{2\pi}}\left[e^{-\frac{(4-2)^2}{2}} + e^{-\frac{(4-5)^2}{2}} + e^{-\frac{(4-7)^2}{2}} + e^{-\frac{(4-8)^2}{2}}\right]$
$= \frac{1}{4\sqrt{2\pi}}\left[e^{-2} + e^{-\frac{1}{2}} + e^{-\frac{9}{2}} + e^{-8}\right]$
$= \mathbf{0.0751318}$
When $x = 4$, $h = 2$, $p_2(x = 4) = \mathbf{0.0971931}$
When $x = 4$, $h = 3$, $p_2(x = 4) = \mathbf{0.0919010}$
(b) When $x = 6$, $h = 1$, $p_1(x = 6) = \mathbf{0.134517}$
When $x = 6$, $h = 2$, $p_2(x = 6) = \mathbf{0.1250115}$
When $x = 6$, $h = 3$, $p_3(x = 6) = \mathbf{0.1031854}$
(c) For $h = 1$, $\log likelihood\ of\ validation\ set$
$= \log p_1(x = 4) + \log p_1(x = 6)$
$= \log 0.0751318 + \log 0.134517$
$= -4.59458$
For $h = 2$, $\log likelihood\ of\ validation\ set$
$= \log p_2(x = 4) + \log p_2(x = 6)$
$= -4.41041$
For $h = 3$, $\log likelihood\ of\ validation\ set$
$= \log p_3(x = 4) + \log p_3(x = 6)$
$= -4.6583$
Since $h = 2$ results in the highest log likelihood in the validation set, therefore $h = 2$ is the best (out of the possibilities of 1, 2 and 3). In practice, one would perform a more exhaustive search over values of h.

Tut 2: Given same data samples as Qn 1, what is the 2-NN estimate of p(x) at x = 3?
Two closest neighbors to $x = 3$ → pt 2 and 5. In this case, $V = 2 \times (5 - 3)$

Tut 3(a): Given the training data $x_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, x_4 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ with corresponding labels $y_1 = 0, y_2 = 0, y_3 = 1, y_4 = 1$, What is the 3NN classification of $x = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$

For $x = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$, 3 closest neighbours are x_1, x_2, x_4



$p(y = 0|x) = \frac{2}{3}$ → no. in class 0 / 3 samples so divide by 3
$(y = 1|x) = \frac{1}{3}$ → no. in class 1 / 3 samples so divide by 3
MAP estimate is class 0 (take largest p)

Tut 3 qn 1: Let $x_1, \cdots, x_N$ be independent samples from a Gaussian distribution. The Gaussian distribution has a known mean of μ and an unknown variance σ^2. Find the ML estimate of σ^2.

$\sigma^2_{ML} = argmax_{\sigma^2}\, p(x_1,\ldots,x_N|\sigma^2)$ → $p(observation|estimate)$
$= argmax_{\sigma^2} \prod_{n=1}^N p(x_1,\ldots,x_N|\sigma^2)$ → by conditional independence
$= argmax_{\sigma^2} \log \prod_{n=1}^N p(x_1,\ldots,x_N|\sigma^2)$ → log is monotonic
$= argmax_{\sigma^2} \sum_{n=1}^N \log p(x_n|\sigma^2)$
$= argmax_{\sigma^2} \sum_{n=1}^N \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\right)$ → $p(x_n|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$
$= argmax_{\sigma^2} \sum_{n=1}^N \left(-\log\sqrt{2\pi\sigma^2} - \frac{(x_n-\mu)^2}{2\sigma^2}\log e\right)$
$= argmax_{\sigma^2} \sum_{n=1}^N \left(-\log\sigma - \frac{(x_n-\mu)^2}{2\sigma^2}\right)$ → drop constants
$= argmax_{\sigma^2}\left(-N\log\sigma - \sum_{n=1}^N \frac{(x_n-\mu)^2}{2\sigma^2}\right)$
$\frac{\partial}{\partial\sigma}\left(-N\log\sigma - \sum_{n=1}^N \frac{(x_n-\mu)^2}{2\sigma^2}\right) = -N\left(\frac{1}{\sigma}\right) - \sum_{n=1}^N \frac{(x_n-\mu)^2}{2}(-2)\sigma^{-3}$ → differentiate
$= -\frac{N}{\sigma} + \sum_{n=1}^N \frac{(x_n-\mu)^2}{\sigma^3}$
$-\frac{N}{\sigma} + \sum_{n=1}^N \frac{(x_n-\mu)^2}{\sigma^3} = 0$ → equate to zero
$\sum_{n=1}^N \frac{(x_n-\mu)^2}{\sigma^3} = \frac{N}{\sigma}$
$\sum_{n=1}^N (x_n - \mu)^2 = N\sigma^2$ → $\boxed{\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^N (x_n - \mu)^2}$

Tut 3 qn 2: Consider a Poisson distribution $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$. Suppose we observe N independent samples from the Poisson distribution: $D = \{x_1,\ldots,x_N\}$. What is the maximum likelihood (ML) estimate of λ?

$\lambda_{ML} = argmax_\lambda p(x_1,\ldots,x_N|\lambda)$ → $p(observation|estimate)$
$= argmax_\lambda \prod_{n=1}^N p(x_1,\ldots,x_N|\lambda)$ → by conditional independence
$= argmax_\lambda \log\prod_{n=1}^N p(x_1,\ldots,x_N|\lambda)$ → log is monotonic
$= argmax_\lambda \sum_{n=1}^N \log p(x_n|\lambda)$
$= argmax_\lambda \sum_{n=1}^N \log\left(\frac{\lambda^{x_n}}{x_n!}e^{-\lambda}\right)$ → substitute Poisson eq
$= argmax_\lambda \sum_{n=1}^N \left(x_n\log\lambda + \log\frac{1}{x_n!} - \lambda\log e\right)$
$= argmax_\lambda \left((\log\lambda)(\sum_{n=1}^N x_n) - N\lambda\right)$ → drop constants
$\frac{\partial}{\partial\lambda}\left((\log\lambda)(\sum_{n=1}^N x_n) - N\lambda\right) = \frac{1}{\lambda}(\sum_{n=1}^N x_n) - N$ → differentiate
$\frac{1}{\lambda}(\sum_{n=1}^N x_n) - N = 0$ → equate to zero
$\frac{1}{\lambda}(\sum_{n=1}^N x_n) = N$ → $\boxed{\lambda_{ML} = \frac{1}{N}\sum_{n=1}^N x_n}$

Tut 3 qn 4: Let x_n be a random walk defined as $x_0 = 0$, $x_N = x_0 + \sum_{n=1}^N z_n$, where z_n is an i.i.d. process with $p(z_n = -1) = p(z_n = 1) = \frac{1}{2}$. Define the absolute value random process $y_n = |x_n|$. Find $p(\max_{1\le n\le 20} y_n = 10|y_{20} = 0)$.
When $y_{20} = 0$, $x_{20} = 0$
When $y_n = 10$, $x_n = 10$ or $x_n = -10$
Since x_n is even, x_0 must be even.
$p(\max_{1\le n\le 20} y_n = 10|y_{20} = 0) = \frac{p(\max_{1\le n\le 20} y_n = 10, y_{20} = 0)}{p(y_{20})}$
$= \frac{2}{\binom{20}{10}}$ → only 2 possibilities; → $p(y_{20}|x_0 = 0)$ → $\frac{(n+k)}{2} = \frac{20+0}{2} = 10$

Tut 3 qn 5: Let x_n be a random walk defined as $p(x_0) = \begin{cases} \frac{1}{5}, & x_0 \in \{-2,-1,0,1,2\} \\ 0, & otherwise \end{cases}$ and $x_N = x_0 + \sum_{n=1}^N z_n$, where z_n is an i.i.d. process with $p(z_n = -1) = p(z_n = 1) = \frac{1}{2}$. What is $p(x_0|x_{11} = 2)$?
Let $x_n = k$: **When x_0 is even:** $(n + k)$ must be even
When x_0 is odd: $(n + k)$ must be odd
Given that $x_{11} = 2$, $(n + k)$ must be odd. → x_0 is odd (only left $x_0 = -1$ or $x_0 = 1$)
Hence, $p(x_0 = -2|x_{11} = 2) = p(x_0 = 0|x_{11} = 2) = p(x_0 = 2|x_{11} = 2) = 0$
$p(x_0 = -1|x_{11} = 2) = \frac{p(x_0=-1)p(x_{11}=2|x_0=-1)}{p(x_{11}=2)} = \frac{\frac{1}{5}\binom{11}{7}\left(\frac{1}{2^{(11)}}\right)}{p(x_{11}=2)}$ → $n + k = \frac{11+2}{2} = 6.5$
$p(x_0 = 1|x_{11} = 2) = \frac{p(x_0=1)p(x_{11}=2|x_0=1)}{p(x_{11}=2)} = \frac{\frac{1}{5}\binom{11}{6}\left(\frac{1}{2^{(11)}}\right)}{p(x_{11}=2)}$ →
$p(x_{11} = 2) = p(x_0 = -1)p(x_{11} = 2|x_0 = -1) + p(x_0 = 1)p(x_{11} = 2|x_0 = 1)$
$$p(x_0|x_{11} = 2) = \begin{cases} \frac{\frac{1}{5}\binom{11}{7}\left(\frac{1}{2^{(11)}}\right)}{\frac{1}{5}\binom{11}{7}\left(\frac{1}{2^{(11)}}\right) + \frac{1}{5}\binom{11}{6}\left(\frac{1}{2^{(11)}}\right)} \\ \frac{\frac{1}{5}\binom{11}{6}\left(\frac{1}{2^{(11)}}\right)}{\frac{1}{5}\binom{11}{7}\left(\frac{1}{2^{(11)}}\right) + \frac{1}{5}\binom{11}{6}\left(\frac{1}{2^{(11)}}\right)} \end{cases} = \begin{cases} \frac{5}{12}, & when\ x_0 = -1 \\ \frac{7}{12}, & when\ x_0 = 1 \\ 0, & otherwise \end{cases}$$

Tut 3 qn 6: Let x_n be a random walk defined as $x_0 = 0$, $x_N = x_0 + \sum_{n=1}^N z_n$, where $z_n, n \ge 0$ is a discrete time white Gaussian noise process, i.e..., $z_1, z_2, \ldots$ are i.i.d $N(0,1)$. Given $x_1 = 4$, $x_2 = 2$, and $0 \le x_3 \le 4$, find the MMSE estimate of x_4.
$x_4^{MMSE} = E_{p(x_4|x_1=4,x_2=2,0\le x_3\le 4)}(x_4)$
$= E_{p(x_4|x_1=4,x_2=2,-2\le z_3\le 2)}(x_4)$ → $x_3 = x_2 + z_3$ → $z_3 = x_3 - x_2$
→$p(a|b) = p(a_1|b) + p(a_2|b)$ → $x_4 = x_3 + z_4 = x_2 + z_3 + z_4$
$= E_{p(x_2|x_1=4,x_2=2,-2\le z_3\le 2)}(x_2) +$ → $\delta(x_2 = 2) = 2$
$E_{p(z_3|x_1=4,x_2=2,-2\le z_3\le 2)}(z_3) +$ → $p(z_3| -2 \le z_3 \le 2)$ is symmetric about 0
$E_{p(z_4|x_1=4,x_2=2,-2\le z_3\le 2)}(z_4)$ → $z_4 = x_3 + z_4$ (independent of x_1, x_2, z_3) → 0
$= 2$

21/22: Consider a uniform distribution $p(x) = U(0,a) = \begin{cases} \frac{1}{a}, 0 \le x \le a \\ 0, otherwise \end{cases}$. Suppose we observe data $D = \{x_1,\ldots,x_n\}$. What is the maximum likelihood estimate of a?
$p(x_1,\ldots,x_n|a) = \left(\frac{1}{a}\right)^n \times \delta(a \ge \max(D))$
$p(x|a) = \begin{cases} 0, if\ a < \max(D) \\ 1, otherwise \end{cases}$
a must be as small as possible but $a \ge \max(x_1,\ldots,x_n)$ → $\boxed{a_{ML} = \max(x_1,\ldots x_n)}$