Chap1. Probability 53+6*8

1.1 Fundamental

- geometric series $\frac{a(1-r^N)}{1-r} = a + ar + ar^2 + ar^3 + ... + ar^{N-1}$
- binomial theorem $(x+y)^N = \sum_{n=0}^{N} C_n^N x^n y^{N-n}$
- Taylor series

$$df = \frac{f_x}{1!}(dx) + \frac{f_{xx}}{2!}(dx)^2 + \frac{f_{xxx}}{3!}(dx)^3 + ... \qquad \text{univariate}$$

$$df = \left[\begin{array}{l} \frac{f_x}{1!}(dx) + \frac{f_y}{1!}(dy) + \\ \frac{f_{xx}}{2!}(dx)^2 + 2\frac{f_{xy}}{2!}(dx)(dy) + \frac{f_{yy}}{2!}(dy)^2 + \\ \frac{f_{xxx}}{3!}(dx)^3 + 3\frac{f_{xxy}}{3!}(dx)^2(dy) + 3\frac{f_{xyy}}{3!}(dx)(dy)^2 + \frac{f_{yyy}}{3!}(dy)^3 + ... \end{array}\right. \qquad \text{bivariate}$$

$$\Delta f = J_f(\Delta X) + \frac{1}{2!}(\Delta x)^T H(\Delta x) + ... \quad where \ J_f = [f_{x_1}, f_{x_2}, ..., f_{x_M}] \qquad \text{multivariate}$$

$$\Delta F = J_F(\Delta X) + ... \quad where \ J_F = \begin{bmatrix} J_{f1} \\ J_{f2} \\ ... \\ J_{fN} \end{bmatrix} \qquad \text{multifunction}$$

- exponential $e^x = \sum_{n=0}^{N} \frac{x^n}{n!}$
- Poisson $\Pr(n \mid \lambda t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$

1.2 Exponential

- definition of exp $e^x = \sum_{n=0}^{N} \frac{x^n}{n!}$
- definition of exp $e^x = \lim_{N\to\infty} \left(1 + \frac{x}{N}\right)^N$
- usage of exp
- continuous discounting / continuous compounding
- differentiation of exponential returns exponential
- convolution of exponential returns exponential
- product of exponential returns exponential
- Fourier of exponential returns exponential
- Euler identity, which is useful to model cycles

1.3 Distributions

	Gaussian	*Bernoulli*	*Binomial*	*Geometric*	*Poisson*	*Exponential*
pdf	$\Pr(X = x \mid \mu\sigma)$	$\Pr(\aleph = n \mid p)$	$\Pr(\aleph = n \mid Np)$	$\Pr(\aleph = n \mid p)$	$\Pr(\aleph = n \mid \lambda T)$	$\Pr(T = t \mid \lambda)$
mean	μ	p	Np	$1/p$	λT	$1/\lambda$
var	σ^2	pq	Npq	$(1/p)\times(q/p)$	λT	$(1/\lambda)^2$
char-fct	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
sum	*yes*		*yes*		*yes*	
min	*yes*			*yes*		*yes*
memoryless				*yes*		*yes*
comparison						*yes*
trick for mean			C_n^N	d/dp	$N!$	*integration by parts*
trick for var			$E[N(N-1)]$	$E[N(N-1)]$	$E[N(N-1)]$	*integration by parts*

Gaussian sum $Gaussian(\mu_1\sigma_1) + Gaussian(\mu_2\sigma_2) = Gaussian(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

Binomial sum $Binomial(N_1 p) + Binomial(N_2 p) = Binomial(N_1 + N_2, p)$

Poisson sum $Poisson(\lambda_1 T) + Poisson(\lambda_2 T) = Poisson(\lambda_1 + \lambda_2, T)$

Characteristic function of a random variable X

$$
\begin{aligned}
\varphi_X(z) &= E[e^{izX}] && \varphi_X : \Re \to C \\
&= \int_{-\infty}^{+\infty} izxd\,\Pr(X < x) && \textit{where } \Pr(X<x) \textit{ is cumulative distribution function of } X \\
&= \int_{-\infty}^{+\infty} izx\,\Pr(X = x)dx && \textit{where } \Pr(X=x) \textit{ is probability density function of } X
\end{aligned}
$$

Taylor series :

$$
\begin{aligned}
\varphi_X(z) &= E[e^{izX}] \\
&= E\left[e^{iz0} + ize^{iz0}X + \frac{(iz)^2 e^{iz0}X^2}{2!} + \frac{(iz)^3 e^{iz0}X^3}{3!} + \frac{(iz)^4 e^{iz0}X^4}{4!} + \dots\right] \\
&= 1 + izE[X] + \frac{(iz)^2}{2!}E[X^2] + \frac{(iz)^3}{3!}E[X^3] + \frac{(iz)^4}{4!}E[X^4] + \dots \quad \textit{weighted sum of all moments}
\end{aligned}
$$

Identical characteristic functions implies equivalent sets of moments, in turn implies the same distribution :

$$
\varphi_{X_1}(z) = \varphi_{X_2}(z) \quad \leftrightarrow \quad \Pr(X_1 = x) = \Pr(X_2 = x)
$$

1.3a Gaussian

pdf

$$
\Pr(X = X \mid U\Sigma) = \frac{1}{\sqrt{(2\pi)^{rank(\Sigma)}\det(\Sigma)}}\exp(-\frac{1}{2}(X-U)^T\Sigma^{-1}(X-U))
$$

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

mean

$$
\begin{aligned}
E[\varepsilon] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-x^2/2}xdx \\
&= \frac{-1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} de^{-x^2/2} \\
&= \frac{-1}{\sqrt{2\pi}}[e^{-x^2/2}]_{-\infty}^{+\infty} && = \mathit{0} \\
E[X \mid \mu = 0, \sigma = 1] &= E[\varepsilon] && = \mathit{0} \\
E[X \mid \mu, \sigma] &= E[\mu + \sigma\varepsilon] && = \mu
\end{aligned}
$$

var

$$
\begin{aligned}
V[\varepsilon] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-x^2/2}x^2dx \\
&= \frac{-1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} xde^{-x^2/2} \\
&= \frac{-1}{\sqrt{2\pi}}[xe^{-x^2/2}]_{-\infty}^{+\infty} - \frac{-1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-x^2/2}dx \\
&= \frac{-1}{\sqrt{2\pi}}[xe^{-x^2/2}]_{-\infty}^{+\infty} + 1 && = \mathit{1} \qquad \textit{since } \lim_{x\to\pm\infty}\frac{x}{e^{x^2/2}} = \lim_{x\to\pm\infty}\frac{1}{xe^{x^2/2}} = 0 \\
V[X \mid \mu = 0, \sigma = 1] &= V[\varepsilon] && = \mathit{1} \\
V[X \mid \mu, \sigma] &= V[\mu + \sigma\varepsilon] = \sigma^2 V[\varepsilon] && = \sigma^2
\end{aligned}
$$

char-fct

$$
\begin{aligned}
\varphi_X(z) &= E[e^{izX}] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2}e^{izx}dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty}\exp(-\frac{x^2 - 2(\mu + i\sigma^2 z)x + \mu^2}{2\sigma^2})dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty}\exp(-\frac{(x-(\mu+i\sigma^2 z))^2 + \mu^2 - (\mu + i\sigma^2 z)^2}{2\sigma^2})dx \\
&= \exp(-\frac{\mu^2 - (\mu + i\sigma^2 z)^2}{2\sigma^2}) \\
&= \exp(\frac{-\mu^2 + \mu^2 + 2i\mu\sigma^2 z + (i\sigma^2 z)^2}{2\sigma^2}) && = \exp(i\mu z - (\sigma z)^2/2)
\end{aligned}
$$

$$
\begin{array}{llll}
\textit{sum} & X_1 & \sim & Gaussian(\mu_1,\sigma_1) \\
 & X_2 & \sim & Gaussian(\mu_2,\sigma_2) \\
 & Y & = & X_1+X_2
\end{array}
$$

$$
\begin{aligned}
\Pr(Y=y\mid X_1X_2) &= \Pr(X_1=y\mid\mu_1\sigma_1)\otimes\Pr(X_2=y\mid\mu_2\sigma_2) \\
\varphi_Y(z) &= \varphi_{X_1}(z)\cdot\varphi_{X_2}(z) \\
&= \exp(i\mu_1 z-(\sigma_1 z)^2/2)\exp(i\mu_2 z-(\sigma_2 z)^2/2) \\
&= \exp(i(\mu_1+\mu_2)z-(\sqrt{\sigma_1^2+\sigma_2^2}\,z)^2/2) \\
X_1+X_2 &\sim Gaussian(\mu_1+\mu_2,\sqrt{\sigma_1^2+\sigma_2^2})
\end{aligned}
$$

$$
\begin{array}{llll}
\textit{min} & X_1 & \sim & Gauss(\mu_1,\sigma_1) \\
 & X_2 & \sim & Gauss(\mu_2,\sigma_2) \\
 & Y & = & \min(X_1,X_2)
\end{array}
$$

$$
\begin{aligned}
\Pr(Y\le y\mid X_1X_2) &= 1-\Pr(Y>y\mid X_1X_2) \\
&= 1-\Pr(X_1>y\mid\mu_1\sigma_1)\Pr(X_2>y\mid\mu_2\sigma_2) \\
&= 1-(1-\Pr(X_1\le y\mid\mu_1\sigma_1))(1-\Pr(X_2\le y\mid\mu_2\sigma_2)) \\
&= \Pr(X_1\le y\mid\mu_1\sigma_1)+\Pr(X_2\le y\mid\mu_2\sigma_2)-\Pr(X_1\le y\mid\mu_1\sigma_1)\Pr(X_2\le y\mid\mu_2\sigma_2) \\
&= \left[\begin{array}{l}
\dfrac{1}{\sqrt{2\pi\sigma_1^2}}\displaystyle\int_{-\infty}^{y} e^{-(x-\mu_1)^2/2\sigma_1^2}\,dx \\
+\dfrac{1}{\sqrt{2\pi\sigma_2^2}}\displaystyle\int_{-\infty}^{y} e^{-(x-\mu_2)^2/2\sigma_2^2}\,dx \\
-\dfrac{1}{2\pi\sigma_1\sigma_2}\displaystyle\int_{-\infty}^{y}\int_{-\infty}^{y} e^{-(x_1-\mu_1)^2/2\sigma_1^2}e^{-(x_2-\mu_2)^2/2\sigma_2^2}\,dx_1dx_2
\end{array}\right. \\
\Pr(Y=y\mid X_1X_2) &= \frac{d}{dy}\Pr(Y\le y\mid X_1X_2) \\
&= \ldots
\end{aligned}
$$

1.3b Bernoulli

$$
\begin{array}{lllllll}
\textit{pdf} & \Pr(X=x\mid p) & = & \left[\begin{array}{lll} p & \textit{if} & x=1 \\ 1-p & \textit{if} & x=0 \end{array}\right. & & & \\
\textit{mean} & E[X\mid p] & = & p\times 1+(1-p)\times 0 & = & p & \\
\textit{var} & V[X\mid p] & = & p\times 1^2+(1-p)\times 0^2-E^2(X) & = & p-p^2 & = \quad pq \\
\textit{char-fct} & \varphi_X(z) & = & E(e^{izX}) & & & \\
 & & = & pe^{iz1}+qe^{iz0} & & & \\
 & & = & 1-p+pe^{iz} & & &
\end{array}
$$

1.3c-f About event waiting

Relationship among various event and waiting distributions :

binomial	=	*#events*	*poisson*	=	*#events per unit time*
geometric	=	*#trials to first success*	*exponential*	=	*time to first success*
negative binomial	=	*#trials to Nth success*	*gamma*	=	*time to Nth success*

Other models include :

- bivariate Poisson, which is sum of two correlated Poissons
- compound Poisson, which is sum of N Poissons

1.3c Binomial

When we perform a sequence of N independent Bernoulli trials, the total number of successful trial is a discrete random variable $\aleph$, with distribution described by Binomial distribution, which is valid within range *[0, N]*.

pdf	$\Pr(\aleph = n \mid Np)$	=	$C_n^N p^n q^{N-n} \quad \forall n \in [0, N]$	
mean	$E[\aleph \mid Np]$	=	$\sum_{n=0}^{N} C_n^N p^n q^{N-n} n$	
		=	$\sum_{n=1}^{N} C_n^N p^n q^{N-n} n$	*since the 1st term is zero*
		=	$\sum_{n=1}^{N} \frac{N!}{(N-n)!n!} p^n q^{N-n} n$	
		=	$Np \times \left[\sum_{n=1}^{N} \frac{(N-1)!}{((N-1)-(n-1))!(n-1)!} p^{n-1} q^{(N-1)-(n-1)} \right]$	
		=	$Np \times \left[\sum_{m=0}^{N-1} \frac{(N-1)!}{((N-1)-m)!m!} p^m q^{(N-1)-m} \right]$	*putting* $m = n-1$
		=	$Np \times (p+q)^{N-1}$	
		=	Np	
var	$E[\aleph(\aleph-1) \mid Np]$	=	$\sum_{n=0}^{N} C_n^N p^n q^{N-n} n(n-1)$	
		=	$\sum_{n=2}^{N} C_n^N p^n q^{N-n} n(n-1)$	*since the 1st and 2nd terms are zero*
		=	$\sum_{n=2}^{N} \frac{N!}{(N-n)!n!} p^n q^{N-n} n(n-1)$	
		=	$N(N-1)p^2 \times \left[\sum_{n=2}^{N} \frac{(N-2)!}{((N-2)-(n-2))!(n-2)!} p^{n-2} q^{(N-2)-(n-2)} \right]$	
		=	$N(N-1)p^2 \times \left[\sum_{m=0}^{N-2} \frac{(N-2)!}{((N-2)-m)!m!} p^m q^{(N-2)-m} \right]$	*putting* $m = n-2$
		=	$N(N-1)p^2 \times (p+q)^{N-2}$	
		=	$N(N-1)p^2$	
	$V[\aleph \mid Np]$	=	$E[\aleph(\aleph-1) \mid Np] + E[\aleph \mid Np] - E^2[\aleph \mid Np]$	
		=	$N(N-1)p^2 + Np - (Np)^2$	
		=	$Np - Np^2$	
		=	Npq	
char-fct	$\varphi_{\aleph}(z)$	=	$E[e^{iz\aleph}]$	
		=	$\sum_{n=0}^{N} C_n^N p^n q^{N-n} e^{izn}$	
		=	$\sum_{n=0}^{N} C_n^N (pe^{iz})^n q^{N-n}$	
		=	$(pe^{iz} + q)^N$	
sum	$\aleph_1$	~	$Bino(N_1, p)$	
	$\aleph_2$	~	$Bino(N_2, p)$	
	$\aleph$	=	$\aleph_1 + \aleph_2$	
	$\Pr(\aleph = n \mid \aleph_1 \aleph_2)$	=	$\Pr(\aleph_1 = n \mid N_1 p) \otimes \Pr(\aleph_2 = n \mid N_2 p)$	
	$\varphi_{\aleph}(z)$	=	$\varphi_{\aleph_1}(z) \cdot \varphi_{\aleph_2}(z)$	
		=	$(pe^{iz} + q)^{N_1} (pe^{iz} + q)^{N_2}$	
		=	$(pe^{iz} + q)^{N_1 + N_2}$	
	$\aleph_1 + \aleph_2$	~	$Binomial(N_1 + N_2, p)$	

1.3d Geometric

Given a sequence of N independent Bernoulli trials, geometric distribution gives the number of trials until the first success.

pdf

$$\Pr(\aleph = n \mid p) = pq^{n-1} \quad \forall n \in [1,\infty)$$

$$\Pr(\aleph > n \mid p) = q^{n} \quad \forall n \in [0,\infty)$$

mean

$$\begin{aligned}
E[\aleph \mid p] &= \textstyle\sum_{n=1}^{\infty} pq^{n-1}n \\
&= p\textstyle\sum_{n=1}^{\infty}(1-p)^{n-1}n \\
&= -p\textstyle\sum_{n=1}^{\infty}\frac{d}{dp}(1-p)^{n} \\
&= -p\frac{d}{dp}\textstyle\sum_{n=1}^{\infty}(1-p)^{n} \\
&= -p\frac{d}{dp}\frac{1-p}{p} \\
&= -p\frac{d}{dp}(1/p-1) \\
&= p/p^{2} \\
&= 1/p
\end{aligned}$$

$$\begin{aligned}
E[\aleph(\aleph-1) \mid p] &= \textstyle\sum_{n=2}^{\infty} pq^{n-1}n(n-1) \\
&= p(1-p)\textstyle\sum_{n=2}^{\infty}(1-p)^{n-2}n(n-1) \\
&= p(1-p)\textstyle\sum_{n=2}^{\infty}\frac{d^{2}}{dp^{2}}(1-p)^{n} \\
&= p(1-p)\frac{d^{2}}{dp^{2}}\textstyle\sum_{n=2}^{\infty}(1-p)^{n} \\
&= p(1-p)\frac{d^{2}}{dp^{2}}\frac{(1-p)^{2}}{p} \\
&= p(1-p)\frac{d}{dp}\left[-\frac{2(1-p)}{p}-\frac{(1-p)^{2}}{p^{2}}\right] \\
&= p(1-p)\frac{d}{dp}\left[-\frac{(1-p)(1+p)}{p^{2}}\right] \\
&= p(1-p)\frac{d}{dp}\left[\frac{p^{2}-1}{p^{2}}\right] \\
&= p(1-p)\frac{d}{dp}(1-1/p^{2}) \\
&= 2p(1-p)/p^{3} \\
&= 2q/p^{2}
\end{aligned}$$

var

$$\begin{aligned}
V[\aleph \mid p] &= E[\aleph(\aleph-1) \mid p] + E[\aleph \mid p] + E^{2}[\aleph \mid p] \\
&= 2q/p^{2} + 1/p - 1/p^{2} \\
&= 2q/p^{2} - q/p^{2} \\
&= q/p^{2}
\end{aligned}$$

char-fct

$$\begin{aligned}
\varphi_{\aleph}(z) &= E[e^{iz\aleph}] \\
&= \textstyle\sum_{n=1}^{\infty} pq^{n-1}e^{izn} \\
&= \textstyle\sum_{n=1}^{\infty} pe^{iz}(qe^{iz})^{n-1} \\
&= \frac{pe^{iz}}{1-qe^{iz}}
\end{aligned}$$

min

$$\begin{aligned}
\aleph_1 &\sim Geom(p_1) \\
\aleph_2 &\sim Geom(p_2) \\
\aleph &= \min(\aleph_1, \aleph_2)
\end{aligned}$$

$$\begin{aligned}
\Pr(\aleph \le n \mid \aleph_1\aleph_2) &= 1-\Pr(\aleph > n \mid \aleph_1\aleph_2) \\
&= 1-q_1^{n}q_2^{n} \\
\Pr(\aleph = n \mid \aleph_1\aleph_2) &= \Pr(\aleph \le n \mid \aleph_1\aleph_2) - \Pr(\aleph \le n-1 \mid \aleph_1\aleph_2) \\
&= 1-q_1^{n}q_2^{n}-1+q_1^{n-1}q_2^{n-1} \\
&= (1-q_1q_2)q_1^{n-1}q_2^{n-1}
\end{aligned}$$

memoryless

$$\begin{aligned}
\Pr(\aleph = n \mid \aleph > n_0, p) &= \frac{\Pr(\aleph = n \mid p)}{\Pr(\aleph > n_0 \mid p)} \\
&= \frac{pq^{n}}{q^{n_0}} \\
&= pq^{n-n_0} \\
&= \Pr(\aleph = n-n_0 \mid p)
\end{aligned}$$

1.3e Poisson

Poisson distribution is a discrete distribution, which counts the number of events occur within a time period t, assuming :

- events occur with a **known** expected event rate λ and
- events occur **independently** of the time since last event occurred.

pdf

$$\Pr(\aleph_{bin} = n \mid Np) = C_n^N p^n q^{N-n} \qquad \textit{where } \aleph_{bin} \textit{ is binomial distributed random variable}$$

$$\begin{aligned}
\Pr(\aleph = n \mid \lambda t) &= \lim_{N\to\infty} \Pr(\aleph_{bin} = n \mid Np) && \textit{where } \aleph \textit{ is Poisson distributed random variable, so that } Np = \lambda t \\
&= \lim_{N\to\infty} C_n^N p^n q^{N-n} \\
&= \lim_{N\to\infty} \frac{N!}{(N-n)!n!}\left(\frac{\lambda t}{N}\right)^n \left(1-\frac{\lambda t}{N}\right)^{N-n} \\
&= \lim_{N\to\infty} \frac{N(N-1)(N-2)...(N-n+1)}{n!}\left(\frac{\lambda t}{N}\right)^n \left(1-\frac{\lambda t}{N}\right)^{-n}\left(1-\frac{\lambda t}{N}\right)^{N} && \textit{we will apply } \lim_{N\to\infty}\left(1-\frac{\lambda t}{N}\right)^N = e^{-\lambda t} \\
&= \lim_{N\to\infty} \frac{N(N-1)(N-2)...(N-n+1)}{n!}\left(\frac{\lambda t}{N-\lambda t}\right)^n \left(1-\frac{\lambda t}{N}\right)^{N} \\
&= \lim_{N\to\infty} \frac{(\lambda t)^n}{n!}\left(1-\frac{\lambda t}{N}\right)^{N} \frac{N}{N-\lambda t}\frac{N-1}{N-\lambda t}\frac{N-2}{N-\lambda t}...\frac{N-n+1}{N-\lambda t} \\
&= \frac{(\lambda t)^n e^{-\lambda t}}{n!} && \textit{since } \lim_{N\to\infty}\frac{N-i}{N-\lambda t} = 1 \textit{ for all } i
\end{aligned}$$

mean

$$\begin{aligned}
E[\aleph \mid \lambda t] &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} n \\
&= \sum_{n=1}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} n && \textit{since } \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \times 0 = 0 \\
&= \lambda t \sum_{n=1}^{\infty} \frac{(\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!} \\
&= \lambda t \sum_{m=0}^{\infty} \frac{(\lambda t)^m e^{-\lambda t}}{m!} \\
&= \lambda t && \textit{since } \sum_{m=0}^{\infty} \frac{(\lambda t)^m e^{-\lambda t}}{m!} = 1
\end{aligned}$$

var

$$\begin{aligned}
E[\aleph^2 \mid \lambda t] &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} n^2 \\
&= \sum_{n=1}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} n^2 && \textit{since } \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \times 0^2 = 0 \\
&= \lambda t \sum_{n=1}^{\infty} \frac{(\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!} n \\
&= \lambda t \sum_{m=0}^{\infty} \frac{(\lambda t)^m e^{-\lambda t}}{m!} (m+1) \\
&= \lambda t \times (E[\aleph \mid \lambda t] + E[1 \mid \lambda t]) \\
&= \lambda t \times (\lambda t + 1) && \textit{since } \sum_{m=0}^{\infty} \frac{(\lambda t)^m e^{-\lambda t}}{m!} = 1
\end{aligned}$$

$$\begin{aligned}
V[\aleph \mid \lambda t] &= E[\aleph^2 \mid \lambda t] - E^2[\aleph \mid \lambda t] \\
&= \lambda t \times (\lambda t + 1) - (\lambda t)^2 \\
&= \lambda t
\end{aligned}$$

char-fct

$$\begin{aligned}
\varphi_{\aleph}(z) &= E[e^{iz\aleph}] \\
&= \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t} e^{izn}}{n!} \\
&= \left(\sum_{n=0}^{\infty} \frac{(\lambda t \exp(iz))^n}{n!}\right) \times e^{-\lambda t} \\
&= e^{\lambda t \exp(iz)} \times e^{-\lambda t} && \textit{since } e^x = \textstyle\sum_{n=0}^{\infty} x^n / n! \\
&= e^{\lambda t(\exp(iz)-1)}
\end{aligned}$$

sum

$$\aleph_k \sim Poisson(\lambda_k, t)$$
$$\aleph = \sum_{k=1}^{K} \aleph_k$$

$$\begin{aligned}
\Pr(\aleph = n \mid \aleph_1 ... \aleph_K) &= \Pr(\aleph_1 = n \mid \lambda_1 t) \otimes \Pr(\aleph_2 = n \mid \lambda_2 t) \otimes ... \otimes \Pr(\aleph_K = n \mid \lambda_K t) \\
\varphi_{\aleph}(z) &= \varphi_{\aleph_1}(z) \cdot \varphi_{\aleph_2}(z) \cdot ... \cdot \varphi_{\aleph_K}(z) \\
&= e^{\lambda_1 t(\exp(iz)-1)} e^{\lambda_2 t(\exp(iz)-1)} ... e^{\lambda_K t(\exp(iz)-1)} \\
&= e^{(\lambda_1+\lambda_2+...+\lambda_K)t(\exp(iz)-1)} \\
\aleph &\sim Poisson(\textstyle\sum_{k=1}^{K} \lambda_k, t)
\end{aligned}$$

2 representations

Since time period t is fixed, some people like to replace λt by λ, resulting in simplier form

$$\begin{aligned}
\Pr(\aleph = n \mid \lambda' = \lambda t) &= \frac{\lambda'^n e^{-\lambda'}}{n!} \\
E[\aleph \mid \lambda' = \lambda t] &= \lambda' \\
V[\aleph \mid \lambda' = \lambda t] &= \lambda' \\
\varphi_{\aleph}(z) &= e^{\lambda'(\exp(iz)-1)}
\end{aligned}$$

shooting star question

Suppose the probability of seeing one or more shooting stars in an hour is *0.44*, what is that probability in half an hour?

$$\begin{aligned}
\Pr(\aleph = 0 \mid \lambda t = c) &= \left.\frac{(\lambda t)^n e^{-\lambda t}}{n!}\right|_{n=0,\lambda t=c} = 1-0.44 \\
& \quad e^{-c} = 0.56 \\
\Pr(\aleph = 0 \mid \lambda t = c/2) &= e^{-c/2} = \sqrt{0.56} \\
\Pr(\aleph > 0 \mid \lambda t = c/2) &= 1-\sqrt{0.56}
\end{aligned}$$

1.3f Exponential

Exponential distribution models the time to the first event of a sequence of Bernoulli trials.

pdf

$$\begin{aligned}
\Pr(T > t \mid \lambda) &= \left.\frac{(\lambda t)^n e^{-\lambda t}}{n!}\right|_{n=0} = e^{-\lambda t} \\
\Pr(T = t \mid \lambda) &= \partial_t \Pr(T < t \mid \lambda) \\
&= \partial_t (1-\Pr(T > t \mid \lambda)) \\
&= \partial_t (1-e^{-\lambda t}) \\
&= \lambda e^{-\lambda t}
\end{aligned}$$

mean

$$\begin{aligned}
E[T \mid \lambda] &= \int_0^\infty \lambda e^{-\lambda t} t dt \\
&= -\int_0^\infty t de^{-\lambda t} \\
&= -[te^{-\lambda t}]_0^\infty + \int_0^\infty e^{-\lambda t} dt \\
&= -[te^{-\lambda t}]_0^\infty - [e^{-\lambda t}]_0^\infty / \lambda \qquad \textit{see remark in LHospital Rule} \\
&= 1/\lambda
\end{aligned}$$

var

$$\begin{aligned}
E[T^2 \mid \lambda] &= \int_0^\infty \lambda e^{-\lambda t} t^2 dt \\
&= -\int_0^\infty t^2 de^{-\lambda t} \\
&= -[t^2 e^{-\lambda t}]_0^\infty + 2\int_0^\infty te^{-\lambda t} dt \qquad \textit{see remark in LHospital Rule} \\
&= 2/\lambda^2 \\
V[T \mid \lambda] &= E(T^2) - E(T)^2 \\
&= 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2
\end{aligned}$$

L Hospital Rule

$$
\begin{array}{llclclclcl}
\textit{Remark 1} & [e^{-\lambda t}]_0^\infty & = & 1/e^{\lambda\infty} - 1/e^{\lambda 0} & = & 0-1 & = & -1 & \Rightarrow & [e^{-\lambda t}]_0^\infty = -1 \\
\textit{Remark 2} & \lim\limits_{t\to\infty} te^{-\lambda t} & = & \lim\limits_{t\to\infty} t/e^{\lambda t} & = & \lim\limits_{t\to\infty} 1/(\lambda e^{\lambda t}) & = & 0 & & \\
 & \lim\limits_{t\to 0} te^{-\lambda t} & = & 0/e^{\lambda 0} & = & 0/1 & = & 0 & \Rightarrow & [te^{-\lambda t}]_0^\infty = 0 \\
\textit{Remark 3} & \lim\limits_{t\to\infty} t^2e^{-\lambda t} & = & \lim\limits_{t\to\infty} t^2/e^{\lambda t} & = & \lim\limits_{t\to\infty} 2t/(\lambda e^{\lambda t}) & = & \lim\limits_{t\to\infty} 2/(\lambda^2 e^{\lambda t}) = & & 0 \\
 & \lim\limits_{t\to 0} t^2e^{-\lambda t} & = & 0^2/e^{\lambda 0} & = & 0/1 & = & 0 & \Rightarrow & [t^2e^{-\lambda t}]_0^\infty = 0 \\
\textit{Remark 4} & E_\lambda(t) & = & 1/\lambda & & & & & & \\
 & \int_0^\infty \lambda e^{-\lambda t} tdt & = & 1/\lambda & & & & & &
\end{array}
$$

$$
\begin{array}{llcl}
\textit{char-fct} & \varphi_T(z) & = & E[e^{izT}] \\
 & & = & \int_0^\infty \lambda e^{-\lambda t} e^{izt} dt \\
 & & = & \int_0^\infty \lambda e^{(-\lambda+iz)t} dt \\
 & & = & \dfrac{\lambda}{-\lambda+iz}[e^{(-\lambda+iz)t}]_0^\infty \\
 & & = & \dfrac{\lambda}{-\lambda+iz}[e^{izt}/e^{\lambda t}]_0^\infty \\
 & & = & \dfrac{1}{1-iz/\lambda}
\end{array}
$$

why? see below

complex number $\cos(zt)+i\sin(zt)$ *lies on the unit circle of Argand diagram for all values of* t *:*

$$
\begin{array}{lcl}
\lim\limits_{t\to\infty} e^{izt}/e^{\lambda t} & = & \lim\limits_{t\to\infty} \cos(zt)/e^{\lambda t} + i \lim\limits_{t\to\infty} \sin(zt)/e^{\lambda t} \quad = \quad 0 \\
\lim\limits_{t\to 0} e^{izt}/e^{\lambda t} & = & e^0/e^0 \quad = \quad 1 \\
{[e^{izt}/e^{\lambda t}]_0^\infty} & = & 0-1 \quad = \quad -1
\end{array}
$$

$$
\begin{array}{llcll}
\textit{sum} & T_k & \sim & \textit{Exponential}(\lambda_k) & \\
 & T & = & \sum_{k=1}^K T_k & \\
 & \Pr(T=t\,|\,T_1...T_K) & = & \Pr(T_1=t\,|\,\lambda_1)\otimes\Pr(T_2=t\,|\,\lambda_2)\otimes...\otimes\Pr(T_K=t\,|\,\lambda_K) & \\
 & \varphi_T(z) & = & \varphi_{T_1}(z)\cdot\varphi_{T_2}(z)\cdot...\cdot\varphi_{T_K}(z) & \\
 & & = & \dfrac{1}{1-iz/\lambda_1}\dfrac{1}{1-iz/\lambda_2}...\dfrac{1}{1-iz/\lambda_K} & \\
 & T & \sim & \textit{Gamma}(K,\lambda) & \textit{if } \lambda_1=\lambda_2=...=\lambda_K
\end{array}
$$

$$
\begin{array}{llcl}
\textit{min and max} & T_k & \sim & \textit{Exponential}(\lambda_k) \quad \forall k\in[1,K] \\
 & T_{\min} & = & \min\limits_{k\in[1,K]} T_k \\
 & T_{\max} & = & \max\limits_{k\in[1,K]} T_k
\end{array}
$$

$$
\begin{array}{lcllcl}
\Pr(T_{\min}>t) & = & \Pr(T_1>t,T_2>t,...,T_K>t) & \Pr(T_{\max}\le t) & = & \Pr(T_1\le t,T_2\le t,...,T_K\le t) \\
 & = & \Pr(T_1>t)\Pr(T_2>t)...\Pr(T_K>t) & & = & \Pr(T_1\le t)\Pr(T_2\le t)...\Pr(T_K\le t) \\
 & = & \prod_{k=1}^K e^{-\lambda_k t} & & = & \prod_{k=1}^K (1-e^{-\lambda_k t}) \\
 & = & e^{-(\sum_{k=1}^K \lambda_k)t} & & = & ... \\
 & = & e^{-\lambda t} \qquad \textit{where } \lambda=\sum_{k=1,K}\lambda_k & & & \\
\Pr(T_{\min}=t) & = & \partial_t \Pr(T_{\min}<t) & \Pr(T_{\max}=t) & = & \partial_t \Pr(T_{\max}\le t) \\
 & = & \partial_t (1-e^{-\lambda t}) & & = & \partial_t \prod_{k=1}^K (1-e^{-\lambda_k t}) \\
 & = & \lambda e^{-\lambda t} & & = & ...
\end{array}
$$

memoryless

$$
\begin{aligned}
\Pr(T > t+s \mid T > s) &= \frac{\Pr(T > t+s)}{\Pr(T > s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t} \\
&= \Pr(T > t) \\
\Pr(T = t+s \mid T > s) &= \Pr(T = t)
\end{aligned}
$$

Comparison between two exponentials

$$
\begin{aligned}
\Pr(T_1 < T_2) &= \int_{t=0}^{t=\infty} d\,\Pr(T_1 < t, T_2 = t) \\
&= \int_{t=0}^{t=\infty} \Pr(T_1 < t)d\,\Pr(T_2 = t) \\
&= \int_{t=0}^{t=\infty} \Pr(T_1 < t)\Pr(T_2 = t)dt \\
&= \int_{t=0}^{t=\infty} (1 - e^{-\lambda_1 t})(\lambda_2 e^{-\lambda_2 t})dt \\
&= \int_{t=0}^{t=\infty} (\lambda_2 e^{-\lambda_2 t} - \lambda_2 e^{-(\lambda_1+\lambda_2)t})dt \\
&= [-e^{-\lambda_2 t}]_0^\infty - [-\lambda_2 /(\lambda_1 + \lambda_2) \times e^{-(\lambda_1+\lambda_2)t}]_0^\infty \\
&= 1 - \lambda_2 /(\lambda_1 + \lambda_2) \\
&= \lambda_1 /(\lambda_1 + \lambda_2)
\end{aligned}
$$

Chap2. Linear Algebra 66 + 21112(10)3

Types of matrix
- symmetric
- triangular
- diagonal
- linear independent
- orthogonal
- orthonormal

	decomposition	*shape of A*	*shape of output*
rank	$A = WR$	*A is rectangular*	*R is row linear independent*
	$A = CW$		*C is column linear independent*
LU	$A = LU$	*A is square*	*L/U are lower/upper triangular*
	$PA = LU$		*P/Q are permutation*
	$PAQ = LU$		
	$A = LDU$		*D is diagonal*
Cholesky	$A = LL^T$	*A is symmatric*	*L/U are lower/upper triangular*
	$A = U^TU$		
LQ/QR	$A = LQ$	*A is rectangular*	*L/U are lower/upper triangular*
	$A = QR$		*Q is orthonormal*
SVD	$A = USV^T$	*A is rectangular*	*U,V are orthonormal, S is diagonal*
eigen	$A = Q\Lambda Q^T$	*A is symmetric*	*Q is orthonormal, Λ is diagonal*

2.1 Rank decomposition 2=[44]

2.1a Meaning of matrix multiplication

Given the following matrices :
- row vector matrix A with size $N\times M$ where N and M are the number and dimension of A
- transformation matrix (or projection matrix) Q with size $M'\times M$, which performs $\mathfrak{R}^M \to \mathfrak{R}^{M'}$
- linear combination matrix (or weight matrix) W with size $N'\times N$, which generates N' vectors from N vectors

then we have 4 interpretations :
1. $N\times N$ dot product of row-vector set A = AA^T
2. $M\times M$ covariance of row-vector set A = $A^T A$
3. $N'\times M'$ weighted sum of transformed A = WAQ^T *Q is transformation matrix in this case*
4. $N\times M$ normalized projection of A on Q = $AQ^T(QQ^T)^{-1}Q$ *Q is projection matrix in this case*

$$\begin{aligned} AQ^T &= \textit{dot product} \\ &= \textit{length of projection} \times \textit{magnitude of } Q \\ AQ^T(QQ^T)^{-1} &= \textit{length of projection} \\ AQ^T(QQ^T)^{-1}Q &= \textit{projection vector of } A \textit{ on } Q \end{aligned}$$

2.1b Span and nullity

1. $span(A) = \{x\}$ s.t. $\exists w, x = wA$ *i.e. linear dependence*
   $null(A) = \{x\}$ s.t. $Ax = 0$ *i.e. zero projection*
2. $A = WR$ *where W is N×K and R is K×M* *i.e. A is decomposed into K linear independent row vectors R*
   $A = CW$ *where C is N×K′ and W is K′×M* *i.e. A is decomposed into K′ linear independent column vectors C*
3. *K is called row rank.*
   K′is called column rank.
   In general, we have : $rank(span(A)) = K = K'$
4. *In general, we have :* $rank(span(A)) + rank(null(A)) = M$

2.2 LU decomposition

LU decomposition is a sequence of elementary operations, including :

- swaping row vector A_{n1} with A_{n2}
- scaling row vector $A_n = scale \times A_n$
- linear operation on row vector $A_{n1} = A_{n1} + scale \times A_{n2}$

$$
\begin{array}{llll}
A & = & L'U & \textit{where } L' \textit{ is lower triangular with all one diagonal} \\
PA & = & L'U & \textit{partial pivoting} \\
PAQ & = & L'U & \textit{full pivoting} \\
A & = & L'DU' & \textit{where } L'U' \textit{ are lower/upper triangular with all one diagonal}
\end{array}
$$

2.3 Cholesky decomposition

Cholesky makes use of symmetric property for decomposition.

$$
\begin{aligned}
A &= LL^T \\
&= \begin{bmatrix} l_{1,1} & 0 & 0 & ... & 0 \\ l_{2,1} & l_{2,2} & 0 & ... & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} & ... & 0 \\ ... & ... & ... & ... & ... \\ l_{N,1} & l_{N,2} & l_{N,3} & ... & l_{N,N} \end{bmatrix} \begin{bmatrix} l_{1,1} & l_{2,1} & l_{3,1} & ... & l_{N,1} \\ 0 & l_{2,2} & l_{3,2} & ... & l_{N,2} \\ 0 & 0 & l_{3,3} & ... & l_{N,3} \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & l_{N,N} \end{bmatrix}
\end{aligned}
$$

Here is the bootstrapping sequence :

$$
\begin{array}{llll}
l_{11} = \sqrt{a_{11}} & & & \\
l_{21} = a_{21}/l_{11} & l_{22} = \sqrt{a_{22} - l_{21}^2} & & \\
l_{31} = a_{31}/l_{11} & l_{32} = (a_{32} - l_{31}l_{21})/l_{22} & l_{33} = \sqrt{a_{33} - l_{31}^2 - l_{32}^2} & \\
l_{41} = a_{41}/l_{11} & l_{42} = (a_{42} - l_{41}l_{21})/l_{22} & l_{43} = (a_{43} - l_{41}l_{31} - l_{42}l_{32})/l_{33} & l_{44} = \sqrt{a_{44} - l_{41}^2 - l_{42}^2 - l_{43}^2}
\end{array}
$$

2.4 LQ/QR decomposition

For every iteration, find the remainder that cannot explaint by existing orthonormal basis.

$$
\begin{aligned}
A &= LQ \\
\begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ ... \\ A_N \end{bmatrix} &= \begin{bmatrix} l_{1,1} & 0 & 0 & ... & 0 \\ l_{2,1} & l_{2,2} & 0 & ... & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} & ... & 0 \\ ... & ... & ... & ... & ... \\ l_{N,1} & l_{N,2} & l_{N,3} & ... & l_{N,N} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ ... \\ Q_N \end{bmatrix} \qquad \textit{where An and Qn are row vectors}
\end{aligned}
$$

$$
\begin{array}{lllll}
Q_1 = \dfrac{A_1}{\|A_1\|} & l_{11} = \|A_1\| & & \\
Q_2 = \dfrac{A_2 - A_2Q_1^TQ_1}{\|A_2 - A_2Q_1^TQ_1\|} & l_{21} = A_2Q_1^T & l_{22} = \|A_2 - A_2Q_1^TQ_1\| & \\
Q_3 = \dfrac{A_3 - A_3Q_1^TQ_1 - A_3Q_2^TQ_2}{\|A_3 - A_3Q_1^TQ_1 - A_3Q_2^TQ_2\|} & l_{31} = A_3Q_1^T & l_{32} = A_3Q_2^T & l_{22} = \|A_3 - A_3Q_1^TQ_1 - A_3Q_2^TQ_2\|
\end{array}
$$

2.5 SVD decomposition

1. $A = USV^T$ *A is N×M, S is N×M, U is N×N, V is M×M*

2. *SVD* by eigen decomposition

$$
\begin{array}{lllll}
AA^T & = (USV^T)(USV^T)^T & = US^2U^T & & \textit{eigen}(AA^T) \textit{ gives U and } S^2 \\
A^TA & = (USV^T)^T(USV^T) & = VS^2V^T & & \textit{eigen}(A^TA) \textit{ gives V and } S^2
\end{array}
$$

2.6 Eigen decomposition

2.6a Three forms of eigen decomposition

Eigen vectors Q are vectors, on applying transformation A gives scaled eigen vectors.

1.	A	=	$Q\Lambda Q^T$	*where Q are row vectors (representation consistent with SVD)*
$\rightarrow$	A^T	=	$Q^T\Lambda Q$	
2.	QA^T	=	ΛQ	*where Q are row vectors (representation consistent with 2.1a3)*
				where A are row vectors of projection matrix (or transformation)

Now, by putting $A^T \rightarrow A$, we have the common representation found in most textbooks :

3.	$Q_n A$	=	λQ_n	*where Q_n is a row vector*
	AQ_n	=	λQ_n	*where Q_n is a column vector*

2.6b Span of eigen vectors

Span of eigen vectors sharing the same eigen value is also an eigen vector of the same eigen value.

$$Q_n A = \lambda Q_n$$
$$Q_m A = \lambda Q_m$$

$$\begin{aligned} \rightarrow \quad (w_n Q_n + w_m Q_m)A &= w_n Q_n A + w_m Q_m A \\ &= w_n \lambda Q_n + w_m \lambda Q_m \\ &= \lambda(w_n Q_n + w_m Q_m) \end{aligned}$$

2.6c When A is special matrix

- When A is identity, any vectors are eigen vectors with eigen value *1*.
- When A is diagonal, any unit vectors δ_n are eigen vectors with eigen value A_{nn}, where δ_n is all zero, but *1* for n^{th} element.
- When A is triangular, eigen values are roots of N degree polynomial :

	$Q_n A$	=	λQ_n	*where Q_n is a row vector*
$\rightarrow$	$Q_n A$	=	$Q_n(\lambda I)$	
$\rightarrow$	0	=	$Q_n(A-\lambda I)$	
$\rightarrow$	0	=	$\det(A-\lambda I)$	*determinant of triangular matrix equals to product of diagonal*
$\rightarrow$	0	=	$\prod_{n=1}^{N}(A_{nn}-\lambda)$	

2.6d When A is symmetric

When A is symmetric, eigen vectors are orthogonal.

$$Q_n A = \lambda_n Q_n \quad \rightarrow \quad Q_n = \lambda_n Q_n A^{-1}$$
$$Q_m A = \lambda_m Q_m \quad \rightarrow \quad Q_m = \frac{1}{\lambda_m} Q_m A$$

$Q_n Q_m^T$	=	$\lambda_n Q_n A^{-1}(\frac{1}{\lambda_m}Q_m A)^T$	
	=	$\frac{\lambda_n}{\lambda_m} Q_n A^{-1} A^T Q_m^T$	
	=	$\frac{\lambda_n}{\lambda_m} Q_n A^{-1} A Q_m^T$	*suppose A is symmetric*
	=	$\frac{\lambda_n}{\lambda_m} Q_n Q_m^T$	
0	=	$\left(1-\frac{\lambda_n}{\lambda_m}\right) Q_n Q_m^T$	
$Q_n Q_m^T$	=	0	*since* $1-\frac{\lambda_n}{\lambda_m} \neq 0$

2.6e When A is semipositive definite

When A is semipositive definite, then all eigen values are positive.

$$\begin{array}{lcll}
X^T AX & \geq & 0 & \forall X \\
X^T (Q\Lambda Q^T) X & \geq & 0 & \forall X \\
(X^T Q)\Lambda(X^T Q)^T & \geq & 0 & \forall X \\
Y\Lambda Y^T & \geq & 0 & \textit{where } Y = X^T Q \\
\prod_{n=1}^{N} \Lambda_{nn} y_n^2 & \geq & 0 & \forall Y \\
\Lambda_{nn} & \geq & 0 & \forall n
\end{array}$$

2.6f Square root matrix and power matrix

- $$\begin{array}{rcllcl}
A & = & Q\Lambda Q^T & & & \\
A^{1/2} A^{1/2} & = & Q\Lambda^{1/2}\Lambda^{1/2}Q^T & & = & (Q\Lambda^{1/2}Q^T)(Q\Lambda^{1/2}Q^T) \\
A^{1/2} & = & Q\Lambda^{1/2}Q^T & & &
\end{array}$$
- $$\begin{array}{rcllcl}
A & = & Q\Lambda Q^T & & & \\
A^2 & = & (Q\Lambda Q^T)(Q\Lambda Q^T) & & = & Q\Lambda^2 Q^T \\
A^3 & = & (Q\Lambda Q^T)(Q\Lambda Q^T)(Q\Lambda Q^T) & & = & Q\Lambda^3 Q^T
\end{array}$$

2.6g Similar matrix

Similar matrices share the same eigen values.

$$\begin{array}{rcll}
B & = & XAX^T & \textit{definition of similarity} \\
 & = & X(Q\Lambda Q^T)X^T & \\
 & = & (XQ)\Lambda(XQ)^T & \textit{hence B has the same eigen values as A}
\end{array}$$

2.6h Implementation 1 : characteristic equation

This method solves for all eigen values.

$$\begin{array}{rcll}
Q_n A & = & \lambda Q_n & \\
Q_n(A-\lambda I) & = & 0 & \\
\det(A-\lambda I) & = & 0 & \rightarrow \textit{solve polynomial forall } \lambda s
\end{array}$$

2.6i Implementation 2 : power method

This method solves for max eigen value only. Given initial guess row vector $V^{(1)}$, we repeatedly apply transformation A on it :

$$\begin{array}{rcllll}
V^{(2)} & = & V^{(1)}A & & & \\
V^{(3)} & = & V^{(2)}A & = & V^{(1)}AA & \\
V^{(4)} & = & V^{(3)}A & = & V^{(1)}AAA & \\
V^{(T+1)} & = & V^{(T)}A & = & V^{(1)}\underbrace{AA...A}_{T} & \textit{now suppose } V^{(1)} = WQ \\
 & & & = & WQ\underbrace{AA...A}_{T} & \\
 & & & = & W\Lambda Q\underbrace{AA...A}_{T-1} \;=\; ... \;=\; W\Lambda^T Q \;=\; \sum_{n=1}^{N} w_n \Lambda_{nn}^T Q_n &
\end{array}$$

$$\frac{[V^{(T+1)}]_m}{[V^{(T)}]_m} = \frac{[\sum_{n=1}^{N} w_n \Lambda_{nn}^T Q_n]_m}{[\sum_{n=1}^{N} w_n \Lambda_{nn}^{T-1} Q_n]_m} \qquad \textit{now suppose } \Lambda_{11}^T > \Lambda_{22}^T > \Lambda_{33}^T$$

$$\lim_{T\to\infty} \frac{[V^{(T+1)}]_m}{[V^{(T)}]_m} = \lim_{T\to\infty} \frac{[\sum_{n=1}^{N} w_n \Lambda_{nn}^T Q_n]_m}{[\sum_{n=1}^{N} w_n \Lambda_{nn}^{T-1} Q_n]_m} = \lim_{T\to\infty} \frac{[w_1 \Lambda_{11}^T Q_1]_m}{[w_1 \Lambda_{11}^{T-1} Q_1]_m} = \Lambda_{11}$$

2.6j Implementation 3 : QR algorithm

This method solves for all eigen vectors. By repeated construction and decomposition :

construction				*decomposition*		
$A^{(1)}$	$=$	A	$\rightarrow$	$A^{(1)}$	$=$	$Q^{(1)}R^{(1)}$
$A^{(2)}$	$=$	$R^{(1)}Q^{(1)}$	$\rightarrow$	$A^{(2)}$	$=$	$Q^{(2)}R^{(2)}$
$A^{(3)}$	$=$	$R^{(2)}Q^{(2)}$	$\rightarrow$	$A^{(3)}$	$=$	$Q^{(3)}R^{(3)}$
$A^{(T+1)}$	$=$	$R^{(T)}Q^{(T)}$				
	$=$	$Q^{(T)T}Q^{(T)}R^{(T)}Q^{(T)}$		*Don't confuse iteration T with transpose.*		
	$=$	$Q^{(T)T}A^{(T)}Q^{(T)}$				
	$=$	$Q^{(T)T}Q^{(T-1)T}A^{(T-1)}Q^{(T-1)}Q^{(T)}$		*Repeat the process on* $A^{(T)}$		
	$=$	$...$				
	$=$	$(Q^{(1)}Q^{(2)}...Q^{(T)})^T A(Q^{(1)}Q^{(2)}...Q^{(T)})$				
	$=$	$Q^T AQ$		*where* $Q = Q^{(1)}Q^{(2)}...Q^{(T)}$		

As T tends to infinity, it is proved that (not shown here) :

$$\lim_{T\to\infty} A^{(T)} = U$$

2.7 Applications

2.7a Solving AX=B

Given $AX = B$, we can solve by :

- *matrix inverse* $\quad A^T AX = A^T B \quad X = (A^T A)^{-1}(A^T B)$
- *LU decomposition* $\quad A^T AX = A^T B \quad LUX = A^T B \quad X = U^{-1}L^{-1}A^T B$
- *Cholesky decomposition* $\quad A^T AX = A^T B \quad LL^T X = A^T B \quad X = (L^T)^{-1}L^{-1}A^T B$
- *LQ decomposition* $\quad AX = B \quad QRX = B \quad X = R^{-1}Q^T B$
- *SVD decomposition* $\quad AX = B \quad USV^T X = B \quad X = VS^{-1}U^T B$
- *eigen decomposition* $\quad A^T AX = A^T B \quad Q\Lambda Q^T X = B \quad X = Q\Lambda^{-1}Q^T B$

$$\begin{bmatrix} l_{1,1} & 0 & 0 & \ldots & 0 \\ l_{2,1} & l_{2,2} & 0 & \ldots & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ l_{M,1} & l_{M,2} & l_{M,3} & \ldots & l_{M,M} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \ldots \\ x_M \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \ldots \\ b_M \end{bmatrix} \qquad \begin{bmatrix} x_1 = b_1 / l_{11} \\ x_2 = (b_2 - l_{21}x_1)/l_{22} \\ x_3 = (b_3 - l_{31}x_1 - l_{32}x_2)/l_{33} \\ \ldots \\ x_M = (b_M - l_{M1}x_1 - l_{M2}x_2 - \ldots)/l_{MM} \end{bmatrix}$$

- *inverse of L and U is bootstrapping*
- *inverse of S and Λ is simply element inverse*
- *inverse of Q is simply transpose*

2.7b Determinant

Given square matrix A, its determinant can be found :

- *Cramer rule*
- *LU decomposition* $\quad \det(A) = \det(LU) = \det(L)\det(U) = \prod_{n=1}^{N} l_{n,n} \prod_{n=1}^{N} u_{n,n}$
- *Cholesky decomposition* $\quad \det(A) = \det(LL^T) = \det(L)\det(L^T) = (\prod_{n=1}^{N} l_{n,n})^2$
- *LQ decomposition* $\quad \det(A) = \det(QR) = \det(Q)\det(R) = \prod_{n=1}^{N} r_{n,n}$
- *SVD decomposition* $\quad \det(A) = \det(USV^T) = \det(U)\det(S)\det(V^T) = \prod_{n=1}^{N} s_{n,n}$
- *eigen decomposition* $\quad \det(A) = \det(Q\Lambda Q^T) = \det Q)\det(\Lambda)\det(Q^T) = \prod_{n=1}^{N} \lambda_n$

2.7c PCA

Find the projection Q such that covariance of AQ is maximized and diagonalized. Here we try to hackle the first objective.

$$\begin{aligned} L &= (AQ)^T AQ - \lambda(Q^T Q - I) && \textit{maximize covariance (after projection) such that } Q^T Q = I \\ &= Q^T (A^T A)Q - \lambda(Q^T Q - I) \\ \frac{\partial L}{\partial Q} &= 0 \\ (A^T A)Q &= \lambda Q \\ Q_{opt} &= \textit{eigenvector}(A^T A) \end{aligned}$$

Comparison

Three different approaches for 2D linear regression :

2.7c PCA as covariance maximization $\quad Q_{opt} = \textit{eigenvector}(A^T A)$

3.2h non-vertical 2D linear regression $\quad \textit{slope} = \dfrac{\mathrm{cov}(x,y)}{\mathrm{var}(x)}$

3.3d PCA as constrainted regression $\quad X_{opt} = \textit{eigenvector}((A - l\overline{A})^T W(A - l\overline{A}))$

Chap3. Linear Regression [333]955

3.1 From likelihood to least square

Let's start with maximum likelihood. Given a set of data points (observations), likelihood is defined as joint probability conditional on a hypothetic parameter set θ.

- $L(\theta) \quad = \quad \Pr(D = d_1, d_2, ..., d_N \mid \theta)$ — *joint probability*
- $L(\theta) \quad = \quad \prod_{n=1}^{N} \Pr(D_n = d_n \mid \theta)$ — *product of probability*
- $\ln L(\theta) \quad = \quad \sum_{n=1}^{N} \ln \Pr(D_n = d_n \mid \theta)$ — *sum of log probability*

We apply linear regression model *AX-B* and Gaussian noise model :

$$\arg\max_{\theta} \ln L = \arg\max_{X} \ln\left[\frac{1}{\sqrt{(2\pi)^{rank(\Sigma)}\det(\Sigma)}} \exp\left(-\frac{1}{2}(AX-B)^T \Sigma^{-1}(AX-B)\right)\right]$$

$$= \arg\max_{X}\left(-\frac{1}{2}(AX-B)^T \Sigma^{-1}(AX-B)\right) \qquad \textit{removing constants, } \Sigma \textit{ is given}$$

$$= \arg\min_{X}(AX-B)^T \Sigma^{-1}(AX-B) \qquad \textit{max problem becomes min problem}$$

We finally convert maximum likelihood to minimum error, i.e. various least square regressions :

- *OLS* — *ordinary least square for uncorrelated-homoskedasticity* — *when* $\Sigma = I$, $\Sigma^{-1} = I$, $\det(\Sigma) = 1$
- *WLS* — *weighted least square for uncorrelated-heteroskedasticity* — *when* $\Sigma = [\sigma_n^2]$, $\Sigma^{-1} = [\sigma_n^{-2}]$, $\det(\Sigma) = \prod_{n=1}^{N}\sigma_n^2$
- *GLS* — *generalized least square for correlated Gaussian*

3.2 Various least square regressions

We will discuss 7 regressions and 2 interpretations :

- *OLS* — $L = (AX-B)^T(AX-B)$
- *WLS* — $L = (AX-B)^T W(AX-B)$
- *GLS* — $L = (AX-B)^T \Sigma^{-1}(AX-B)$
- *Tikhonov regularization* — $L = (AX-B)^T \Sigma^{-1}(AX-B) + (X-X_0)^T \Sigma_Q^{-1}(X-X_0)$ — *given anchor* X_0
- *nonlinear - Gauss Newton* — $L = (F(A \mid X)-B)^T \Sigma^{-1}(F(A \mid X)-B)$ — *given initial guess* X_0
- *nonlinear - Levenberg Marq* — $L = (F(A \mid X)-B)^T \Sigma^{-1}(F(A \mid X)-B) + \lambda(\Delta X)^T(\Delta X)$ — *given initial guess* X_0
- *nonlinear noise model - iterWLS* — $L = \sum_{n=1}^{N} g(A_n X - b_n)$
- *interpretation in 2D* — $slope = \mathrm{cov}(x,y)/\mathrm{var}(x)$
- *interpretation in regressor space* — $AX = proj_A(B)$

$$\textit{where } F(A \mid X) = \begin{bmatrix} f(A_1 \mid X) \\ f(A_2 \mid X) \\ f(A_3 \mid X) \\ ... \\ f(A_N \mid X) \end{bmatrix} \qquad J_F = \begin{bmatrix} \partial f(A_1 \mid X)/\partial x_1 & \partial f(A_1 \mid X)/\partial x_2 & \partial f(A_1 \mid X)/\partial x_3 & ... & \partial\partial f(A_1 \mid X)/\partial x_M \\ \partial f(A_2 \mid X)/\partial x_1 & \partial f(A_2 \mid X)/\partial x_2 & \partial f(A_2 \mid X)/\partial x_3 & ... & \partial f(A_2 \mid X)/\partial x_M \\ \partial f(A_3 \mid X)/\partial x_1 & \partial f(A_3 \mid X)/\partial x_2 & \partial f(A_3 \mid X)/\partial x_3 & ... & \partial f(A_3 \mid X)/\partial x_M \\ ... & ... & ... & ... & ... \\ \partial f(A_N \mid X)/\partial x_1 & \partial f(A_N \mid X)/\partial x_2 & \partial f(A_N \mid X)/\partial x_3 & ... & \partial f(A_N \mid X)/\partial x_M \end{bmatrix}$$

Here are the corresponding solutions :

- *OLS* — $\partial_X L = 2A^T(AX-B) \quad \rightarrow \quad X_{OLS} = (A^T A)^{-1}(A^T B)$
- *WLS* — $\partial_X L = 2A^T W(AX_{WLS}-B) \quad \rightarrow \quad X_{WLS} = (A^T WA)^{-1}(A^T WB)$
- *GLS* — $\partial_X L = 2A^T \Sigma^{-1}(AX_{GLS}-B) \quad \rightarrow \quad X_{GLS} = (A^T \Sigma^{-1} A)^{-1}(A^T \Sigma^{-1} B)$
- *Tikhonov regularization* — $\partial_X L = 2A^T \Sigma^{-1}(AX_{TR}-B) + 2\Sigma_Q^{-1}(X_{TR}-X_0) \quad \rightarrow \quad X_{TR} = (A^T \Sigma^{-1} A + \Sigma_Q^{-1})^{-1}(A^T \Sigma^{-1} B + \Sigma_Q^{-1} X_0)$

- *nonlinear - Gauss Newton*

$$\begin{aligned}
L &= (F(A\,|\,X)-B)^T\Sigma^{-1}(F(A\,|\,X)-B) \\
&= (F(A\,|\,X^{(t)})+J_F\Delta X-B)^T\Sigma^{-1}(F(A\,|\,X^{(t)})+J_F\Delta X-B) \\
&= (J_F\Delta X-(B-F(A\,|\,X^{(t)})))^T\Sigma^{-1}(J_F\Delta X-(B-F(A\,|\,X^{(t)}))) \\
\partial_{\Delta X}L &= 2J_F^T\Sigma^{-1}(J_F\Delta X-(B-F(A\,|\,X^{(t)}))) \\
\Delta X &= (J_F^T\Sigma^{-1}J_F)^{-1}(J_F^T\Sigma^{-1}(B-F(A\,|\,X^{(t)})))
\end{aligned}$$

- *nonlinear - Levenberg Marq*

$$\begin{aligned}
L &= (F(A\,|\,X)-B)^T\Sigma^{-1}(F(A\,|\,X)-B)+\lambda(\Delta X)^T(\Delta X) \\
&= (F(A\,|\,X^{(t)})+J_F\Delta X-B)^T\Sigma^{-1}(F(A\,|\,X^{(t)})+J_F\Delta X-B)+\lambda(\Delta X)^T(\Delta X) \\
&= (J_F\Delta X-(B-F(A\,|\,X^{(t)})))^T\Sigma^{-1}(J_F\Delta X-(B-F(A\,|\,X^{(t)})))+\lambda(\Delta X)^T(\Delta X) \\
\partial_{\Delta X}L &= 2J_F^T\Sigma^{-1}(J_F\Delta X-(B-F(A\,|\,X^{(t)})))+2\lambda I\Delta X \\
\Delta X &= (J_F^T\Sigma^{-1}J_F+\lambda I)^{-1}(J_F^T\Sigma^{-1}(B-F(A\,|\,X^{(t)})))
\end{aligned}$$

- *non Gauss noise model - iterWLS*

$$\begin{aligned}
L &= \textstyle\sum_{n=1}^N g(A_nX-b_n) && \textit{where regression model } AX{=}B \textit{ is linear} \\
\partial_{\Delta X}L &= \sum_{n=1}^N \frac{\partial g(e_n)}{\partial e_n}\frac{\partial e_n}{\partial X} && \textit{where } e_n = A_nX-b_n \\
&= \sum_{n=1}^N w_ne_n\frac{\partial e_n}{\partial X} && \textit{where } w_n = \frac{1}{e_n}\frac{\partial g(e_n)}{\partial e_n} \\
&= \frac{1}{2}\sum_{n=1}^N w_n\frac{\partial e_n{}^2}{\partial X} \\
&= \frac{1}{2}\frac{\partial}{\partial X}\sum_{n=1}^N w_ne_n{}^2
\end{aligned}$$

- *interpretation in 2D, let's consider :*

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_N & 1 \end{bmatrix} \qquad B = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix} \qquad X = \begin{bmatrix} m \\ c \end{bmatrix}$$

$$W = diag(w_1, w_2, ..., w_N)$$

we then have :

$$\begin{aligned}
A^TWA &= \begin{bmatrix} \sum_{n=1}^N w_nx_nx_n & \sum_{n=1}^N w_nx_n \\ \sum_{n=1}^N w_nx_n & \sum_{n=1}^N w_n \end{bmatrix} &&= \begin{bmatrix} S_{wxx} & S_{wx} \\ S_{wx} & S_w \end{bmatrix} \\
A^TWB &= \begin{bmatrix} \sum_{n=1}^N w_nx_ny_n \\ \sum_{n=1}^N w_ny_n \end{bmatrix} &&= \begin{bmatrix} S_{wxy} \\ S_{wy} \end{bmatrix} \\
m &= \frac{S_{wxy}S_w-S_{wx}S_{wy}}{S_{wxx}S_w-S_{wx}S_{wx}} &&= \frac{\mathrm{cov}(x,y)}{\mathrm{var}(x)} \\
c &= \frac{S_{wxx}S_{wy}-S_{wx}S_{wxy}}{S_{wxx}S_w-S_{wx}S_{wx}} &&= \frac{S_{wxx}S_{wy}-S_{wx}S_{wxy}}{\mathrm{var}(x)}
\end{aligned}$$

- *interpretation in regressor space*

AX = *span of column vector A, weighted by X*

then distance between span of A and vector B is minimised when span of A equals to the projection of B on A :

$$\begin{aligned}
AX_{OLS} &= proj_A(B) && \textit{(visualise the case when } N >> M = 2\textit{)} \\
&= A(A^TA)^{-1}A^TB \\
\Rightarrow \quad X_{OLS} &= (A^TA)^{-1}A^TB
\end{aligned}$$

3.3 Various constraints

We will discuss 5 constraints :

- linear constraint 1 $\min_X \|AX - B\|_W^2$ *so that* $X_1 = X_1^*$ *where* $X_1 = column\_matrix(K \times 1)$
- linear constraint 2 $\min_X \|AX - B\|_W^2$ *so that* $X_1 = CX_2 + D$ *where* $X_1 = column\_matrix(K \times 1)$
- linear constraint 3 $\min_X \|AX - B\|_W^2$ *so that* $CX = D$
- quad constraint 1 $\min_{X,y} \|AX - ly\|_W^2$ *so that* $X^T X = 1$ *where* $l = ones(N \times 1)$ *and y is scalar, distance to origin*

$$\rightarrow \lambda X = \text{cov}(A)X \quad \textit{and} \quad y = \overline{A}X = \frac{l^T WAX}{l^T Wl}$$

- quad constraint 2 $\min_X \|AX - B\|_W^2$ *so that* $X^T X = 1$

Here are the corresponding solutions :

- $$\begin{aligned} L &= (AX-B)^T W(AX-B) \\ &= (A_1X_1 + A_2X_2 - B)^T W(A_1X_1 + A_2X_2 - B) \\ &= (A_2X_2 - (B - A_1X_1^*))^T W(A_2X_2 - (B - A_1X_1^*)) \\ 0 &= 2A_2^T W(A_2X_2 - (B - A_1X_1^*)) \qquad \textit{differentiate L wrt } X_2\textit{, set it zero} \\ X_2 &= (A_2^T WA_2)^{-1}(A_2^T W(B - A_1X_1^*)) \end{aligned}$$

- $$\begin{aligned} L &= (AX-B)^T W(AX-B) \\ &= (A_1X_1 + A_2X_2 - B)^T W(A_1X_1 + A_2X_2 - B) \\ &= (A_1(CX_2 + D) + A_2X_2 - B)^T W(A_1(CX_2 + D) + A_2X_2 - B) \\ &= ((A_1C + A_2)X_2 - (B - A_1D))^T W((A_1C + A_2)X_2 - (B - A_1D)) \\ 0 &= 2(A_1C + A_2)^T W((A_1C + A_2)X_2 - (B - A_1D)) \qquad \textit{differentiate L wrt } X_2\textit{, set it zero} \\ X_2 &= ((A_1C + A_2)^T W(A_1C + A_2))^{-1}((A_1C + A_2)^T W(B - A_1D)) \end{aligned}$$

- $$\begin{array}{lllllll} AX & = & B & \textit{so that} & CX & = & D \\ AVV^T X & = & B & \textit{so that} & USV^T X & = & D \\ A'X' & = & B & \textit{so that} & USX' & = & D \quad \textit{where } A' = AV \textit{ and } X' = V^T X \\ & & & & X' & = & S^{-1}U^T D \\ & & & & \begin{bmatrix} X'_1 \\ X'_2 \end{bmatrix} & = & \begin{bmatrix} S_1^{-1} \\ S_2^{-1} \end{bmatrix} U^T D \quad \textit{ignore null space } S_2 \textit{ of constraint} \\ & & & & X'_1 & = & S_1^{-1}U^T D \end{array}$$

$$\begin{array}{l} \arg\min_X (AX-B)^T W(AX-B) \\ s.t. \quad CX = D \end{array} \Rightarrow \begin{array}{l} \arg\min_{X'} (A'X' - B)^T W(A'X' - B) \\ s.t. \quad X'_1 = S_1^{-1}U^T D \end{array}$$

By transforming the problem, we can apply technique in solving linear constraint 1 :

$$\begin{aligned} X'_2 &= (A'_2{}^T WA'_2)^{-1} - (A'_2{}^T W(B - A'_1 X'_1)) \\ X &= VX' \\ &= \begin{bmatrix} VX'_1 \\ VX'_2 \end{bmatrix} \\ &= \begin{bmatrix} V(S_1^{-1}U^T D) \\ V(A'_2{}^T WA'_2)^{-1}(A'_2{}^T W(B - A'_1 X'_1)) \end{bmatrix} \end{aligned}$$

- *Setting up a Lagrangian, firstly take derivative wrt y, secondly substitute result into Lagrangian, finally take derivative wrt X :*

$$L = (AX - ly)^T W(AX - ly) + \lambda(1 - X^T X)$$

$$
\begin{array}{lll}
0 & = & l^T W(AX - ly) \\
y & = & \dfrac{l^T WAX}{l^T Wl} = \overline{A}X \\
L & = & (AX - l\overline{A}X)^T W(AX - l\overline{A}X) + \lambda(1 - X^T X) \\
 & = & ((A - l\overline{A})X)^T W((A - l\overline{A})X) + \lambda(1 - X^T X) \\
0 & = & 2(A - l\overline{A})^T W(A - l\overline{A})X - 2\lambda X \\
\lambda X & = & \mathrm{cov}(A)X
\end{array}
$$

firstly take derivative wrt y

where $l^T Wl = \sum_{n=1}^N w_n = scalar$ *and* $l^T WA = \sum_{n=1}^N w_n A_n$

secondly substitute result into Lagrangian

finally take derivative wrt X

where $\mathrm{cov}(A) = (A - l\overline{A})^T W(A - l\overline{A})$

- *Setting up a Lagrangian, firstly take derivative wrt X, secondly substitute result into constraint, finally solve for multiplier :*

$$L = (AX - B)^T W(AX - B) + \lambda(1 - X^T X)$$

$$
\begin{array}{lll}
0 & = & 2A^T W(AX - B) - 2\lambda X \\
0 & = & (A^T WA - \lambda I)X - (A^T WB) \\
X & = & (A^T WA - \lambda I)^{-1}(A^T WB) \\
1 & = & X^T X \\
 & = & ((A^T WA - \lambda I)^{-1}(A^T WB))^T ((A^T WA - \lambda I)^{-1}(A^T WB)) \\
f(\lambda) & = & ((A^T WA - \lambda I)^{-1}(A^T WB))^T ((A^T WA - \lambda I)^{-1}(A^T WB)) - 1
\end{array}
$$

firstly take derivative wrt X

secondly substitute result into constraint

finally solve for λ using Newton Raphson

3.4 Various applications

concentric circles fitting

$$\begin{bmatrix} 2x_1 & 2y_1 & -\delta_{c_1} \\ 2x_2 & 2y_2 & -\delta_{c_2} \\ 2x_3 & 2y_3 & -\delta_{c_3} \\ \dots & \dots & \dots \\ 2x_N & 2y_N & -\delta_{c_N} \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ r_1^2 - x_c^2 - y_c^2 \\ \dots \\ r_M^2 - x_c^2 - y_c^2 \end{bmatrix} = \begin{bmatrix} x_1^2 + y_1^2 \\ x_2^2 + y_2^2 \\ x_3^2 + y_3^2 \\ \dots \\ x_N^2 + y_N^2 \end{bmatrix}$$

rectangle fitting

$$\begin{bmatrix} x_1 & 1 & 0 & 0 & 0 \\ -y_2 & 0 & 1 & 0 & 0 \\ x_3 & 0 & 0 & 1 & 0 \\ -y_4 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m \\ c_1 \\ mc_2 \\ c_3 \\ mc_4 \end{bmatrix} = \begin{bmatrix} y_1 \\ x_2 \\ y_3 \\ x_4 \end{bmatrix}$$

homo-inspect

$$\begin{bmatrix} (1 - r_{x_1})(1 - r_{y_1}) & (1 - r_{x_1})r_{y_1} & r_{x_1}(1 - r_{y_1}) & r_{x_1}r_{y_1} \\ (1 - r_{x_2})(1 - r_{y_2}) & (1 - r_{x_2})r_{y_2} & r_{x_2}(1 - r_{y_2}) & r_{x_2}r_{y_2} \\ (1 - r_{x_3})(1 - r_{y_3}) & (1 - r_{x_3})r_{y_3} & r_{x_3}(1 - r_{y_3}) & r_{x_3}r_{y_3} \\ \dots & \dots & \dots & \dots \\ (1 - r_{x_N})(1 - r_{y_N}) & (1 - r_{x_N})r_{y_N} & r_{x_N}(1 - r_{y_N}) & r_{x_N}r_{y_N} \end{bmatrix} \begin{bmatrix} a_{00} \\ a_{01} \\ a_{10} \\ a_{11} \end{bmatrix} = \begin{bmatrix} ins(x_1, y_1) \\ ins(x_2, y_2) \\ ins(x_3, y_3) \\ \dots \\ ins(x_N, y_N) \end{bmatrix}$$

template-inspect

$$\begin{bmatrix} lrn_1(x_1, y_1) & lrn_2(x_1, y_1) & \dots & lrn_M(x_1, y_1) & 1 \\ lrn_1(x_2, y_2) & lrn_2(x_2, y_2) & \dots & lrn_M(x_2, y_2) & 1 \\ lrn_1(x_3, y_3) & lrn_2(x_3, y_3) & \dots & lrn_M(x_3, y_3) & 1 \\ \dots & \dots & \dots & \dots. & \dots \\ lrn_1(x_N, y_N) & lrn_2(x_N, y_N) & \dots & lrn_M(x_N, y_N) & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_M \\ b \end{bmatrix} = \begin{bmatrix} ins(x_1, y_1) \\ ins(x_2, y_2) \\ ins(x_3, y_3) \\ \dots \\ ins(x_N, y_N) \end{bmatrix}$$

alignment, given cost function $L = \sum_{n=1}^N [(x_n^{ins} - x_n^{lrnTrans}) \times \cos(\alpha_n^{lrn}) + (y_n^{ins} - y_n^{lrnTrans}) \times \sin(\alpha_n^{lrn})]^2$

$$\begin{bmatrix} x_1^{lrn}\cos(\alpha_1^{lrn}) + y_1^{lrn}\sin(\alpha_1^{lrn}) & -y_1^{lrn}\cos(\alpha_1^{lrn}) + x_1^{lrn}\sin(\alpha_1^{lrn}) & \cos(\alpha_1^{lrn}) & \sin(\alpha_1^{lrn}) \\ x_2^{lrn}\cos(\alpha_2^{lrn}) + y_2^{lrn}\sin(\alpha_2^{lrn}) & -y_2^{lrn}\cos(\alpha_2^{lrn}) + x_2^{lrn}\sin(\alpha_2^{lrn}) & \cos(\alpha_2^{lrn}) & \sin(\alpha_1^{lrn}) \\ \dots & \dots & \dots & \dots \\ x_N^{lrn}\cos(\alpha_N^{lrn}) + y_N^{lrn}\sin(\alpha_N^{lrn}) & -y_N^{lrn}\cos(\alpha_N^{lrn}) + x_N^{lrn}\sin(\alpha_N^{lrn}) & \cos(\alpha_N^{lrn}) & \sin(\alpha_N^{lrn}) \end{bmatrix} \begin{bmatrix} a \\ b \\ \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} x_1^{ins}\cos(\alpha_1^{lrn}) + y_1^{ins}\sin(\alpha_1^{lrn}) \\ x_2^{ins}\cos(\alpha_2^{lrn}) + y_2^{ins}\sin(\alpha_2^{lrn}) \\ \dots \\ x_N^{ins}\cos(\alpha_N^{lrn}) + y_N^{ins}\sin(\alpha_N^{lrn}) \end{bmatrix}$$