Background of Probability Theory and Perspective

February 19, 2014

This course aims to provide an in-depth knowledge and technical background of Probability Theory for students who study in the related fields of STEM (Science, Technology, Engineering, Mathematics) and their applications. *College Calculus*, *Discrete Mathematics*, *Programming Language*, *Linear Algebra* are prerequisites for this course which are summarized as follows and will be reviewed in *the first week only.* A group study of working on homework assignments is encouraged but

□ **No late homework will be collected**

□ **Plagiarism is completely prohibited**

- ▷ Review of Calculus
- ▷ Review of Linear Algebra
- ▷ Review of Fundamental Matlab Programming

(1) Evaluate $\displaystyle\int_0^\infty x^m \frac{1}{\theta} e^{-x/\theta} dx$ for $m = 0, 1, 2$, respectively.

(2) Evaluate $\displaystyle\int_0^\pi x^n \frac{1}{2} \sin(x) dx$ for $n = 0, 1, 2$, respectively.

(3) Show that $limit_{n\to\infty}(1 + \frac{h}{n})^n = e^h$.

(4) Evaluate the following integrals:

(a) $\int_0^{2\pi} sin(kx)sin(mx)dx$, where $k, m \geq 1,\ k \neq m$.
(b) $\int_0^{2\pi} cos(kx)cos(mx)dx$, where $k, m \geq 1,\ k \neq m$.
(c) $\int_0^{2\pi} sin(kx)cos(mx)dx$, where $k, m \geq 1$.
(d) $\int_0^{2\pi} sin(kx)sin(kx)dx$, where $k \geq 1$.
(e) $\int_0^{2\pi} cos(mx)cos(mx)dx$, where $m \geq 1$.

(5) Let $f(x) = e^{-x} - x$. Show that there exists an $x_0 \in [0, 1]$ such that $f(x_0) = 0$.

(6) Let $A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$.

(a) Find the characteristic polynomial of matrix A.
(b) Find the eigenvalues and corresponding eigenvectors of matrix A.
(c) What are the singular values of A?
(d) Can the Cholesky algorithm be applied to decomposing A? Explain.

□ *Some Solutions*

(4) 0 for (a), (b), (c) and π for (d), (e).

(5) $f(0) = e^{-0} - 0 = 1 > 0$ and $f(1) = e^{-1} - 1 < 0$.

(6) $P_A(x) = x^2 + 4x + 3$, $\lambda(A) = \{-1,\ -3\}$, $\sigma(A) = \{3,\ 1\}$,

Gamma Function and Its Properties

(7) Define $\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt$ for $x > 0$ and let $\gamma = \int_0^\infty e^{-x^2}dx$. Then

(a) $\Gamma(x+1) = x\Gamma(x),\ for\ x > 0,\ \ \Gamma(n) = (n-1)!$ if $n \in N$, where $0! \equiv 1$

(b) $\int_0^\infty e^{-x^2}dx = \frac{\sqrt{\pi}}{2}$

(c) $\Gamma(\frac{1}{2}) = 2\gamma = \sqrt{\pi}$

(d) $\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}$

Proof:

(a)

$$\begin{aligned}
\Gamma(x+1) = \int_0^\infty e^{-t}t^x dt &= \int_0^\infty t^x d(-e^{-t}) \\
&= -t^x e^{-t}\ |_0^\infty + \int_0^\infty x e^{-t}t^{x-1}dt \\
&= x\Gamma(x)
\end{aligned}$$

Thus $\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \cdots = (n-1)!$

(b) $\gamma^2 = \int_0^\infty e^{-x^2}dx \int_0^\infty e^{-y^2}dy = \int_0^\infty \int_0^\infty e^{-(x^2+y^2)}dxdy.$

Let

$$\begin{array}{ll} x = r\cos\theta, & 0 \le \theta \le 2\pi \\ y = r\sin\theta, & 0 \le r < \infty \end{array}\ ,\quad J_{r,\theta} = \frac{\partial(x,y)}{\partial(r,\theta)} = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = r$$

Then

$$\begin{aligned}
\gamma^2 &= \int_0^\infty \int_0^\infty e^{-(x^2+y^2)}dxdy \\
&= \int_0^\infty \int_0^{\pi/2} e^{-r^2} J_{r,\theta} drd\theta \\
&= \int_0^\infty \int_0^{2\pi} re^{-r^2} drd\theta \\
&= \frac{\pi}{4}
\end{aligned}$$

Thus $\gamma = \frac{\sqrt{\pi}}{2}$.

(c) $\Gamma(\frac{1}{2}) = \int_0^\infty e^{-t}t^{\frac{1}{2}-1}dt = \int_0^\infty e^{-x^2}x^{-1}dx^2 = 2\int_0^\infty e^{-x^2}dx = 2\gamma = \sqrt{\pi}.$

Solution for A Linear System of Equations

The central problems of *Linear Algebra* is to study the properties of matrices and to investigate the solutions of linear equations.

▷ The following linear system of equations can be written as $A\mathbf{x} = \mathbf{b}$ with the *augmented* matrix $[A \mid \mathbf{b}]$.

$$\begin{array}{ccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & \cdot & \cdot & + & a_{1n}x_n & = & b_1 \\
a_{21}x_1 & + & a_{22}x_2 & + & \cdot & \cdot & + & a_{2n}x_n & = & b_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & & \\
a_{m1}x_1 & + & a_{m2}x_2 & + & \cdot & \cdot & + & a_{mn}x_n & = & b_m
\end{array}$$

▷ *Example*

$$\begin{array}{ccccccc}
2x & + & y & + & z & = & 5 \\
4x & - & 6y & & & = & -2 \\
-2x & + & 7y & + & 2z & = & 9
\end{array}, \quad
\left[\begin{array}{ccc|c}
2 & 1 & 1 & 5 \\
4 & 6 & 0 & -2 \\
-2 & 7 & 2 & 9
\end{array}\right]$$

- *Matlab Solution*

```
>> A=[2, 1, 1; 4, -6, 0; -2, 7, 2];      % Input matrix A
>> b=[5, -2, 9]';                         % Input vector b
>> x=A\b                                  % Solve for Ax=b
```

Linear Least Squares Problem

Consider the problem of determining an $\mathbf{x} \in R^n$ such that the *residual sum of squares* $\rho^2(\mathbf{x}) = \|\mathbf{b} - A\mathbf{x}\|_2^2$ is minimized for given $\mathbf{b} \in R^n$, $A \in R^{m\times n}$, $m \geq n$.

□ *Example: A Best Line Fit*

Given $[x_i, y_i]^t \in R^2$ for $1 \leq i \leq n$, find a line which best fits these points. The problem is equivalent to finding m and b to minimize

$$f(m, b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2$$

or to solve

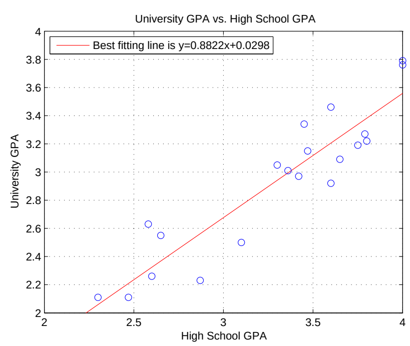
$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \\ \sum_{i=1}^n x_i & \sum_{i=1}^n 1 \end{bmatrix} \begin{bmatrix} m \\ \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \\ \sum_{i=1}^n y_i \end{bmatrix}$$

◇ *dataGPA.txt*

```
20 pairs of High School and University GPAs for Line Fit
3.75   3.19
3.45   3.34
2.87   2.23
3.60   3.46
3.42   2.97
4.00   3.79
2.65   2.55
3.10   2.50
3.47   3.15
2.60   2.26
4.00   3.76
2.30   2.11
2.47   2.11
3.36   3.01
3.60   2.92
3.65   3.09
3.30   3.05
2.58   2.63
3.80   3.22
3.79   3.27
```

Matlab Solution for A Line Fit Problem

```
%
% Script file:  linefit.m
% A Linear Least Square Fit for (GPA_high school, GPA_university)
%
fin=fopen('dataGPA.txt','r');
fgetl(fin);
m=2;  n=20;
T=fscanf(fin,'%f',[m n]);
fclose(fin);
T=T';
X=T(:,1);
Y=T(:,2);
A=[sum(X.*X), sum(X);  sum(X), n];
b=[sum(X.*Y); sum(Y)];
v=A\b;  % (0.8822, 0.0298)
for j=1:n,
  t=2.0+0.2*j;
  X1(j)=t;
  Y1(j)=t*v(1)+v(2);
end
plot(X1,Y1,'r-',X,Y,'bo'); axis([2 4  2 4]); grid;
legend('Best fitting line is y=0.8822x+0.0298','Location','NorthWest')
title('University GPA vs. High School GPA')
ylabel('University GPA')
xlabel('High School GPA')
```



Eigenvalues and Eigenvectors

Let $A \in R^{n\times n}$. If $\exists\ \mathbf{v} \neq \mathbf{0}$ such that $A\mathbf{v} = \lambda\mathbf{v}$, λ is called an eigenvalue of matrix A, and $\mathbf{v}$ is called an eigenvector corresponding to (or belonging to) the eigenvalue λ. Note that $\mathbf{v}$ is an eigenvector implies that $\alpha\mathbf{v}$ is also an eigenvector for all $\alpha \neq 0$. We define the Eigenspace(λ) as the vector space spanned by all of the eigenvectors corresponding to the eigenvalue λ.

Examples:

1. $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, $\lambda_1 = 2$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\lambda_2 = 1$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
2. $A = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$, $\lambda_1 = 2$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\lambda_2 = 1$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.
3. $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$, $\lambda_1 = 4$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\lambda_2 = 2$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.
4. $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $\lambda_1 = j$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ j \end{bmatrix}$, $\lambda_2 = -j$, $\mathbf{u}_2 = \begin{bmatrix} j \\ 1 \end{bmatrix}$, $j = \sqrt{-1}$.
5. $B = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$, then $\lambda_1 = 3$, $\mathbf{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$; $\lambda_2 = -1$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
6. $C = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$, then $\tau_1 = 4$, $\mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$; $\tau_2 = 2$, $\mathbf{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

- $A\mathbf{x} = \lambda\mathbf{x} \ \Rightarrow\ (\lambda I - A)\mathbf{x} = \mathbf{0},\ \mathbf{x} \neq \mathbf{0} \ \Rightarrow\ det(\lambda I - A) = P(\lambda) = 0.$

Computing Eigenvalues and Eigenvectors

Note that $\|\mathbf{u}_i\|_2 = 1$ and $\|\mathbf{v}_i\|_2 = 1$ for $i = 1, 2$. Denote $U = [\mathbf{u}_1, \mathbf{u}_2]$ and $V = [\mathbf{v}_1, \mathbf{v}_2]$, then

$$U^{-1}BU = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}, \quad V^{-1}CV = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$$

$\triangleright$ Note that $V^t = V^{-1}$ but $U^t \neq U^{-1}$.

- $\sum_{i=1}^n \lambda_i \;=\; \sum_{i=1}^n a_{ii} \;=\; tr(A)$
- $\prod_{i=1}^n \lambda_i \;=\; det(A)$

$\diamond$ Let $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

$\triangleright$ Matlab Solution for Computing Eigenvalues and Eigenvectors

```
>>A=[2, -1, 0; -1, 2, -1; 0, -1, 2];
>>[U, D]=eig(A);        % UD=AU
>>norm(A-U*D*U',2)      % Check diagonalization of A
```

Mean, Variance, and Moment Function of Discrete Distributions

Bernoulli $f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$

$M(t) = 1 - p + pe^t; \qquad \mu = p, \ \sigma^2 = p(1-p)$

Binomial $f(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n$

$b(n, p) \qquad M(t) = (1 - p + pe^t)^n; \qquad \mu = np, \ \sigma^2 = np(1-p)$

Geometric $f(x) = (1-p)^{x-1}p, \quad x = 1, 2, \ldots$

$M(t) = \frac{pe^t}{1-(1-p)e^t}, \quad t < -ln(1-p)$

$\mu = \frac{1}{p}, \ \sigma^2 = \frac{1-p}{p^2}$

Negative Binomial $f(x) = C(x-1, r-1)p^r(1-p)^{x-r}, \quad x = r, r+1, r+2, \ldots$

$M(t) = \frac{(pe^t)^r}{[1-(1-p)e^t]^r} \quad t < -ln(1-p)$

$\mu = \frac{r}{p}, \ \sigma^2 = \frac{r(1-p)}{p^2}$

Poisson $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$

$M(t) = e^{\lambda(e^t-1)}; \qquad \mu = \lambda, \ \sigma^2 = \lambda$

Uniform $f(x) = \frac{1}{m}, \quad x = 1, 2, \ldots$

$M(t) = \frac{1}{m} \cdot \frac{e^t(1-e^{mt})}{1-e^t}; \qquad \mu = \frac{m+1}{2}, \ \sigma^2 = \frac{m^2-1}{12}$

Continuous Distributions

Uniform $U(a,b)$ $f(x) = \frac{1}{b-a}$, $a \leq x \leq b$

Exponential $f(x) = \frac{1}{\theta} e^{-x/\theta}$, $0 < x < \infty$

Gamma $f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$, $0 < x < \infty$

$\chi^2(r)$ **Chi-Square** $f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{(r/2)-1} e^{-x/2}$, $0 < x < \infty$

Beta (×) $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0 < x < 1$

$N(\mu, \sigma^2)$ **Normal** $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$, $-\infty < x < \infty$

Gamma, Exponential, χ^2 Distributions

Consider an (approximate) *Posisson distribution* with mean (arrival rate) λ, let the r.v. X be the waiting time until the αth arrival occurs. Then the cumulative distribution of X can be expressed as

$$\begin{aligned} F(x) &= P(X \leq x) = 1 - P(X > x) \\ &= 1 - P(fewer\ than\ \alpha\ arrivals\ in\ (0, x]) \\ &= 1 - \textstyle\sum_{k=0}^{\alpha-1}[e^{-\lambda x}(\lambda x)^k/(k!)] \end{aligned} \tag{1}$$

$$f(x) = F'(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, 0 < x < \infty, \quad \alpha > 0 \tag{2}$$

Let $\theta = 1/\lambda$, we have the p.d.f. of Gamma Distribution

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, 0 < x < \infty \tag{3}$$

For Gamma distribution, if $\alpha = 1$, we have the p.d.f. of Exponential Distribution

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 < x < \infty \tag{4}$$

For Gamma distribution, if $\theta = 2$ and $\alpha = r/2$, where r is a positive integer, then we have the p.d.f. of $\chi^2(r)$ distribution with r degrees of freedom.

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{(r/2)-1} e^{-x/2}, \quad 0 < x < \infty \tag{5}$$

For Gamma Distribution,

$M(t) = 1/(1 - \theta t)^\alpha$, $\mu = \alpha\theta$, $\sigma^2 = \alpha\theta^2$

For Exponential Distribution,

$M(t) = 1/(1 - \theta t)$, $\mu = \theta$, $\sigma^2 = \theta^2$

For $\chi^2(r)$ Distribution,

$M(t) = 1/(1 - 2t)^{r/2}$, $\mu = r$, $\sigma^2 = 2r$

Normal (Gaussian) Distributions

A normal distribution of r.v. $X \sim N(\mu, \sigma^2)$ has the p.d.f.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \tag{6}$$

♣ $M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

♣ $(X - \mu)/\sigma \sim N(0, 1)$

♣ $Z \sim N(0, 1) \;\Rightarrow\; Z^2 \sim \chi^2(1)$.

When $\mu = 0$, $\sigma = 1$, $X \sim N(0, 1)$ is said to have the *standard normal distribution.* The cumulative distribution is denoted as

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad -\infty < z < \infty \tag{7}$$