Subject,
Year. Month. Date. ( )

[redacted]

Kimia Noorbakhsh

Homework 1

1.1) $\mathbb{E}[\mathbb{E}[X|Y]] \stackrel{?}{=} \mathbb{E}[X]$

proof: $\mathbb{E}[\mathbb{E}[X|Y]] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y=y)\, f_Y(y)\, dy$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, x\, dx\, f_Y(y)\, dy$$

$(f_{X|Y}(x|y)\, f_Y(y) = f_{X,Y}(x,y))$ $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\, dx\, dy$

(take x outside) $= \int_{-\infty}^{\infty} x \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy}_{(\text{marginalize out})\ f_X(x)}\, dx = \mathbb{E}[X]$ ✓

$$Var(X) = \mathbb{E}(Var(X|Y)) + Var(\mathbb{E}[X|Y])$$

Proof: by definition: $Var(x) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

from the problem above: $= \mathbb{E}(\mathbb{E}(X^2|Y)) - E(E(X|Y))^2$

$\left(\begin{array}{l} Var(X|Y) = \mathbb{E}(X^2|Y) - \mathbb{E}(X|Y)^2 \\ \downarrow \\ \text{conditional definition of variance} \end{array}\right)$ $= \mathbb{E}(Var(X|Y) + \mathbb{E}[X|Y)^2) - \mathbb{E}(\mathbb{E}(X|Y))^2$

$$= \mathbb{E}(Var(X|Y)) + \underbrace{\mathbb{E}(\mathbb{E}(X|Y)^2) - \mathbb{E}(\mathbb{E}(X|Y))^2}_{\text{by definition: } = Var(\mathbb{E}(X|Y))}$$

$$= \mathbb{E}(Var(X|Y)) + Var(\mathbb{E}(X|Y))$$ ✓

Subject:
Year. Month. Date. ( )

1.2) Let $x_1, x_2, \ldots, x_n \overset{iid}{\sim} \underset{pdf}{f}, \underset{cdf}{F}$

pdf of $\min(x_1, \ldots, x_n) = ?$ $\rightarrow Y_2 = \min(X_1, \ldots, X_n)$

$\Rightarrow F_{Y_2}(y_2) = P(Y_2 \leqslant y_2) = 1 - P(Y_2 > y_2)$

(because Y_2 is min) $= 1 - P(X_1 > y_2, X_2 > y_2, \ldots, X_n > y_2)$

independant $= 1 - P(X_1 > y_2) P(X_2 > y_2) \cdots P(X_n > y_2)$

iid $= 1 - P(X_1 > y_2)^n = 1 - (1 - F_X(y_2))^n$

pdf of $Y_2 = f_{Y_2}(y_2) = \frac{d}{dy_2} F_{Y_2}(y_2) = n(1 - F_X(y_2))^{n-1} f(y_2)$

$\Rightarrow f_{Y_2}(y_2) = n(1 - F_X(y_2))^{n-1} \times f(y_2)$

pdf of $\max(X_1, \ldots, X_n) = ?$ $Y_1 = \max(X_1, \ldots, X_n)$

$F_{Y_1}(y_1) = P(Y_1 \leqslant y_1) = P(X_1 \leqslant y_1, X_2 \leqslant y_1, \ldots, X_n \leqslant y_1)$

independant $= P(X_1 \leqslant y_1) \cdots P(X_n \leqslant y_1)$

iid $= P(X_1 \leqslant y_1)^n = F_X(y_1)^n$

pdf of $Y_1 = f_{Y_1}(y_1) = \frac{d}{dy_1} F_{Y_1}(y_1) = n F_X(y_1)^{n-1} \times f(y_1)$

$\Rightarrow f_{Y_1}(y_1) = n F_X(y_1)^{n-1} \times f(y_1)$

1.3

$P(X=K)=P(Y=K)=pq^k \quad K=0,1,2,\dots$ $\Rightarrow$ $X-Y$ and $\min(X,Y)$ are independent.

$q=1-p$

X, Y are geometric distribution $\Rightarrow P(X>K) = 1-(1-q^{K+1})$
(or $Y>K$) $= q^{K+1}$

if $U=\min(X,Y) \Rightarrow P(U=K) = P(X=k, Y=K) + 2P(X=K, Y>K)$

(x,y independent) $= (pq^K)^2 + 2(pq^K \cdot q^{K+1}) = p^2q^{2K} + 2pq^{2K+1}$

$= q^{2K}(p^2+2pq)$ ①

if $W=X-Y \Rightarrow P(W=K) = \sum_{i=0}^{\infty}(pq^{K+i})(pq^i)$

$= \sum_{i=0}^{\infty} p^2q^Kq^{2i} = p^2q^K \sum_{i=0}^{\infty} q^{2i}$

(geometric series) $= p^2q^K \times \frac{1}{1-q^2} = \frac{p^2q^K}{1-q^2}$

②

now we have : $P(U=K_1, W=K_2)$ $(K_1, K_2 > 0)$
$(X-Y=K_2)$
$Y=\min$ $= P(X=K_1+K_2, Y=K_1) = pq^{K_1} \cdot pq^{K_1+K_2}$ ③

on the other hand: $P(U=K_1)P(W=K_2) \overset{①,②}{=} (p^2+2pq)q^{2K_1} \frac{p^2q^{K_2}}{1-q^2}$

$= \cancel{p}(\underset{1-q}{p}+\underset{1+q}{2q})\, q^{2K_1} \frac{p^2q^{K_2}}{1-q^2} = q^{2K_1}p^2q^{K_2} = ③$

$\Rightarrow P(U=K_1, W=K_2) = P(U=K_1)P(W=K_2)$

$\Rightarrow$ U and W are independent

1.41 0.55 : we design the procedure as follows:

consider all kinds of tossing the coin 5 times in a row. we know that we have 32 conditions. name these conditions $c_1, c_2, \ldots, c_{32}$ consider $c_1, c_2, \ldots, c_{11}$ as win states and consider $c_{12}, c_{13}, \ldots, c_{20}$ as lose states and consider $c_{21}, c_{23}, \ldots, c_{32}$ as try again state. Therefore the probability to win is equal to

$$p(\text{win}) = \underbrace{\frac{11}{32}}_{\text{win first}} + \underbrace{\frac{12}{32}}_{\text{try again}} \times \underbrace{\frac{11}{32}}_{\text{win}} + \underbrace{\left(\frac{12}{32}\right)^2}_{\text{2 try again}} \times \underbrace{\frac{11}{32}}_{\text{then win}} + \cdots$$

$$= \frac{11}{32}\left(1 + \frac{12}{32} + \left(\frac{12}{32}\right)^2 + \cdots\right) = \frac{11}{32} \times \frac{1}{\frac{20}{32}}$$

$$= \frac{11}{32} \times \frac{32}{20} = \frac{11}{20} = 0.55 \checkmark$$

$\frac{1}{3}$: like the above procedure, concider all kinds of tossing the coin 2 times in a row : $\underbrace{(H,H)}_{\frac{1}{4}}, \underbrace{(H,T), (T,H)}_{\frac{1}{2}}, \underbrace{(T,T)}_{\frac{1}{4}}$

consider (H,H) a win, (H,T) and (T,H) a lose and (T,T) → try again

now : $p(\text{win}) \underset{\text{as above}}{\overset{\text{same}}{=}} \frac{1}{4} + \frac{1}{4} \times \frac{1}{4} + \left(\frac{1}{4}\right)^2 \times \frac{1}{4} + \cdots$

$$= \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \checkmark$$

Subject:
Year. Month. Date. ( )

2.1 suppose that a and x are $n\times 1$ vectors and A is $n\times n$ matrix

Suppose that $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

and $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$

$\frac{d}{dx} a^T x \overset{?}{=} a^T$: $\frac{d}{dx} a^T x = \frac{d}{dx} [a_1 \dots a_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \frac{d}{dx} (a_1 x_1 + \dots + a_n x_n)$

$$= \left[\frac{d}{dx_1} (a_1 x_1 + \dots + a_n x_n) \dots \frac{d}{dx_n} (a_1 x_1 + \dots + a_n x_n) \right]$$

$$= [a_1 a_2 \dots a_n] = a^T \checkmark$$

$\frac{d}{dx} x^T A x \overset{?}{=} x^T (A + A^T)$: $x^T A x = [x_1 \dots x_n] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

$$= [(a_{11} x_1 + \dots + a_{n1} x_n) \dots (a_{1n} x_1 + \dots + a_{nn} x_n)] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \left[\sum_{i=1}^{n} a_{i1} x_i \dots \sum_{i=1}^{n} a_{in} x_i \right] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 \sum_{i=1} a_{i1} x_i + \dots + x_n \sum_{i=1}^{n} a_{in} x_i$$

$$= \sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{ij} x_i = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$$

$$\Rightarrow \frac{d}{dx} x^T A x = \left[\frac{d}{dx_1} (x^T A x) \dots \frac{d}{dx_n} (x^T A x) \right]$$

$$\frac{d}{dx_k} (x^T A x) = \frac{d}{dx_k} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j \right)$$

P4PCO

continued

Subject:
Year. Month. Date. ( )

2.1 - continued

$$\Rightarrow \frac{d}{dx_k}(x^TAx) = \frac{d}{dx_k}\left(x_1\sum_{i=1}^{n} a_{i1}x_i + \cdots + x_k\sum_{i=1}^{n} a_{ik}x_i + \cdots + x_n\sum_{i=1}^{n} a_{in}x_i\right)$$

$$= x_1a_{k1} + x_2a_{k2} + \cdots + \left(\sum_{i=1}^{n} a_{ik}x_i + x_ka_{kk}\right) + \cdots + x_na_{kn}$$

$$= \underbrace{\sum_{j=1}^{n} a_{kj}x_j}_{k^{th}\text{ row of }A} + \underbrace{\sum_{i=1}^{n} a_{ik}x_i}_{k^{th}\text{ column of }A = k^{th}\text{ row of }A^T}$$

$$\Rightarrow \frac{d}{dx_k}(x^TAx) = \boxed{x^T(A + A^T)} \checkmark$$

$$\frac{d}{dx}x^TA \stackrel{?}{=} A^T : \quad x^TA = [x_1 \dots x_n]\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} x^TA_1 \\ x^TA_2 \\ \vdots \\ x^TA_n \end{bmatrix}$$

name the columns $\rightarrow$ $A_1 \quad A_n$

$\Rightarrow$ from that we know $\frac{d}{dx}x^Ta = a^T$

$$\Rightarrow \frac{d}{dx}x^TA = \frac{d}{dx}\begin{bmatrix} x^TA_1 \\ \vdots \\ x^TA_n \end{bmatrix} = \begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_n^T \end{bmatrix} = \boxed{A^T} \checkmark$$

2.2 ① determinant is the product of the eigen values:

if A is $n \times n$, let $\lambda_1, \dots, \lambda_n$ are the eigenvalues.

$\Rightarrow \lambda_s$ are roots of the characteristic polynomial:

$$\det(A - \lambda I) = \text{characteristic polynomial}$$

$$= (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)$$
$$= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$$

set $\lambda = 0 \Rightarrow \det(A) = \lambda_1 \lambda_2 \dots \lambda_n$ ✓

---

trace is sum of eigen values:

$$p(\lambda) = \det(A - \lambda I) = \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix}$$

using cofactor expansion

$$= (a_{11} - \lambda)(a_{22} - \lambda) \dots (a_{nn} - \lambda) + b_1 \lambda^{n-2} + b_2 \lambda^{n-3} + \dots + b_n$$

$$= (-\lambda)^n + (-\lambda)^{n-1} [a_{11} + \dots + a_{nn}] + \dots$$

$$\Rightarrow \det(A - \lambda I) = (-1)^n (\lambda^n - \lambda^{n-1} (\operatorname{Tr} A) + \dots)$$

(from above) $= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$

$\rightarrow$ by equality of coefficients $\Rightarrow \lambda_1 + \dots + \lambda_n = \operatorname{Tr} A$ ✓

$\operatorname{eig}(A) = \operatorname{eig}(A^T)$: we know that $\det(A) = \det(A^T)$

now: $(A - \lambda I)^T = A^T - \lambda I^T = A^T - \lambda I \Rightarrow \det(A - \lambda I) = \det(A^T - \lambda I)$

$\Rightarrow A, A^T$ have same characteristic polinomials

$\Rightarrow \operatorname{eig}(A) = \operatorname{eig}(A^T)$ ✓

Subject:
Year. Month. Date. ( )

2.2 continued: 2 $A \in L(V) \to V \overset{?}{=} null(A) \oplus range(A)$

proof: $v_1, v_2, \dots, v_n \to$ basis of V with respect to A (diagnal)

$\Rightarrow$ for every $1 \leqslant i \leqslant n \to \exists \lambda_i \in F : Av_i = \lambda_i v_i$

* without loss of generality, suppose that $\begin{cases} \lambda_j = 0 & \forall 1 \leqslant j \leqslant m \\ \lambda_j \neq 0 & \forall m+1 \leqslant j \leqslant n \end{cases}$

$\Rightarrow V = span\{v_1, \dots, v_m\} \oplus span\{v_{m+1}, \dots, v_n\}$ 1

claim: $span\{v_1, \dots, v_m\} = null(A)$ and $range(A) = \{v_{m+1}, \dots, v_n\}$

proof of claim: $span\{v_1, \dots, v_m\}$ is the eigenspace of A with respect to $0 \Rightarrow \forall v_i \in \{v_1, \dots, v_m\} \to Av_i = 0$

*

also for any $v \in null(A) \to Av = 0$

v is a eigenvector of A with respect to $0 \Rightarrow v \in \{v_1, \dots, v_m\}$

**

*, ** $\Rightarrow null(A) = \{v_1, \dots, v_m\}$ 2

we know that $Av_i = \lambda_i v_i \to A(\lambda_i^{-1} v_i) = v_i \quad \forall m+1 \leqslant i \leqslant n$

$\Rightarrow span\{v_{m+1}, \dots, v_n\} \subseteq range(A)$ 3

also for every $v \in range(A) \to v = A(v'), v' \in V$

because v_is are basis we have: $v' = \sum a_i v_i$, a_is $\in F$

$\Rightarrow v = Av' = A(a_1 v_1 + \dots + a_n v_n) =$

PAPCO

continued

Subject:
Year. Month. Date. ( )

2.2 continued ②

$$\Rightarrow v = Av' = A(a_1v_1 + \dots + a_nv_n) = a_1.Av_1 + \dots + a_n.Av_n$$

$$\begin{bmatrix} (Av_1, \dots, Av_m = 0) \\ Av_j = \lambda_j v_j \\ m+1 \leq j \leq n \end{bmatrix} = a_{m+1}\lambda_{m+1}v_{m+1} + \dots + a_n\lambda_n v_n \in \text{span}\{v_{m+1}, \dots, v_n\}$$

$$\Rightarrow \underbrace{\text{range}(A) \subseteq \text{Span}\{v_{m+1}, \dots, v_n\}}_{④}$$

$$③,④ \Rightarrow \underbrace{\text{range}(A) = \text{span}\{v_{m+1}, \dots, v_n\}}_{⑤}$$

$$①,②,⑤ \Rightarrow \underbrace{V = \text{null}(A) + \text{range}(A)}$$ ✓

2.3 suppose $A = U\Sigma V^T \Rightarrow QR = U\Sigma V^T$

Q orthogonal $\Rightarrow Q^{-1} = Q^T \xrightarrow{\times Q^{-1}} R = Q^{-1}U\Sigma V^T$

$$\Rightarrow \underbrace{R = Q^TU\Sigma V^T}$$

now notice that: $(Q^TU)^TQ^TU = U^TQQ^TU = U^TU = I$

$\Rightarrow Q^TU$ is orthogonal too.

$\Rightarrow R = (Q^TU)\Sigma V^T$ is a SVD decomposition for R.

different from A

Same as A

Same as A

3. $x_1, x_2, \ldots, x_n \overset{iid}{\sim} N(\underset{?}{\mu}, \underset{\text{known}}{\sigma^2})$

- MLE of μ :

$$\text{likelihood} = p(x_1, \ldots, x_n \mid \mu) \overset{iid}{=} p(x_1 \mid \mu) \ldots p(x_n \mid \mu)$$

$$\overset{iid}{=} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\log(\text{likelihood}) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{d}{d\mu} \log(\text{likelihood}) = \sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2}$$

$$\rightarrow \text{set } \frac{d}{d\mu} = 0 \Rightarrow \sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2} = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \mu = n\mu$$

$$\Rightarrow \hat{\mu}_{mle} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

- $\mu \sim N(\nu, \beta^2) \rightarrow$ MAP of μ?

Bayes rule : $p(\mu \mid x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n \mid \mu)\, p(\mu)}{p(x_1, \ldots, x_n)}$ $\circledast$

what we know : $p(\mu) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(\frac{(\mu-\nu)^2}{2\beta^2}\right)$

and

$$p(x_1, \ldots, x_n \mid \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

← continued

Subject:
Year. Month. Date. ( )

3 continued

$\Rightarrow$ from ⊛ we have:

$$\text{posterior} = p(\mu \mid x_1, \ldots, x_n) = \frac{\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right) \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(\frac{(\mu-\nu)^2}{2\beta^2}\right)}{P(x_1, \ldots, x_n) = \text{Const}}$$

$$\Rightarrow \log(\text{posterior}) = \left(\sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x_i-\mu)^2}{2\sigma^2}\right) + \log\left(\frac{1}{2\pi\beta^2}\right) - \frac{(\mu-\nu)^2}{2\beta^2}$$

$$\Rightarrow \frac{\partial}{\partial \mu} \log(\text{posterior}) = \left(\sum_{i=1}^{n} \frac{x_i-\mu}{\sigma^2}\right) - \frac{\mu-\nu}{\beta^2}$$

$$\text{set } \frac{\partial}{\partial \mu} = 0 \Rightarrow \left(\sum_{i=1}^{n} \frac{x_i-\mu}{\sigma^2}\right) - \frac{\mu-\nu}{\beta^2} = 0$$

$$\Rightarrow \frac{\mu-\nu}{\beta^2} = \frac{\sum_{i=1}^{n} x_i}{\sigma^2} - \frac{n\mu}{\sigma^2}$$

$$\Rightarrow \frac{\mu}{\beta^2} + \frac{n\mu}{\sigma^2} = \frac{\sum x_i}{\sigma^2} + \frac{\nu}{\beta^2}$$

$$\mu\left(\frac{\sigma^2+n\beta^2}{\sigma^2\beta^2}\right) = \frac{\beta^2\sum x_i + \nu\sigma^2}{\sigma^2\beta^2}$$

$$\Rightarrow \boxed{\hat{\mu}_{MAP} = \frac{\sigma^2\nu + \beta^2\sum_{i=1}^{n} x_i}{\sigma^2+n\beta^2}}$$

- $N^{(n)} \to \infty$

$$\text{note that } \hat{\mu}_{MAP} = \frac{\sigma^2\nu+\beta^2\sum x_i}{\sigma^2+n\beta^2} = \frac{\sigma^2\nu}{\sigma^2+n\beta^2} + \frac{\beta^2\sum x_i}{\sigma^2+n\beta^2}$$

$$= \frac{\sigma^2\nu}{\sigma^2+n\beta^2} + \frac{\frac{1}{n}\sum x_i}{1+\frac{\sigma^2}{n\beta^2}}$$

PAPCO

continued

Subject:
Year. Month. Date. ( )

3 continued

$$\Rightarrow \hat{\mu}_{MLE} = \frac{\sigma^2 \nu}{n\beta^2+\sigma^2} + \frac{\frac{1}{n}\sum x_i}{1+\frac{\sigma^2}{n\beta^2}}$$

So when $n \to \infty \Rightarrow \frac{\sigma^2}{n\beta^2} \to 0, \quad \frac{\sigma^2 \nu}{n\beta^2+\sigma^2} \to 0$

$$\Rightarrow \hat{\mu}_{map} \longrightarrow \frac{\sum x_i}{n} = \hat{\mu}_{MLE}$$

So when $n \to \infty$ $\hat{\mu}_{MLE}$ and $\hat{\mu}_{map}$ become closer to eachother. ✓