Optimization Theory and Algorithm **Lecture 4 - 2/27/2025**

Lecture 4 Random Vector and Exceptation

Lecturer:Xiangyu Chang *Scribe: Hongyi Xu,Ye Zhang*

Edited by: Zhihong Liu

1 Recall

1.1 CDF

- $F_{X,Y}(x,y) = P(x \leq x, y \leq y)$

1.2 PMF and PDF

- PMF:
  - $f_{X,Y}(x,y) = P(x \leq x, y \leq y)$
- PDF:
  - $f_{X,Y}(x,y) \geq 0$
  - $\int_{\mathbb{R}^2} f_{X,Y}(x,y)\, dx\, dy = 1$
  - $P((X,Y) \in A) = \int_A f_{X,Y}(x,y)\, dx\, dy$

1.3 Marginal Distribution

- $f_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y f_{X,Y}(x,y)$
- $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy$

1.4 Independent

- $P(X \in A, Y \in B) = P(X \in A)P(X \in B)$

2 Conditional PMF and PDF

- $f_{Y|X}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ or $f_{X,Y}(x,y) = f_Y(y|x) \cdot f_X(x)$

if independent:

- $f_Y(y|x) = f_Y(y)$

For example,

- $X \sim U[0,1], Y|X \sim U[x,1], f_Y?$

Step1:

- $f_x(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$
- $f_{Y|X}(y) = \begin{cases} \frac{1}{1-x}, & x \leq y < 1 \\ 0, & \text{otherwise} \end{cases}$

Step2:

- $f_{X,Y}(x,y) = f_Y(y|x) \cdot f_X(x) = \begin{cases} \frac{1}{1-x}, & x \leq y < 1 \\ 0, & \text{otherwise} \end{cases}$

Step3:

- $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx = \int_0^y \frac{1}{1-x}dx = -\ln(1-y)$

supplement:Linear Model

- $\{(x_i, y_i)\}_{i=1}^n$
- $Y_i = \beta^T x_i + \varepsilon_i$

3 Multivarable

3.1 CDF

- $\overrightarrow{x} = x = (x_1, \cdots, x_d)^\top$
  $F_Y(y) = P(y_1 \leq x_1, \cdots, y_d \leq x_d)$

3.2 PMF and PDF

- $F_Y(y) = P(y_1 = x_1, \cdots, y_d = x_d)$

- $\int_{R^d} f_x(x) = 1$

- $P(x \in A) = \int_A f_x(x)dx, A \subseteq R^d$

- $f_{x_1 \cdots x_d}(x_1, \cdots, x_k) = \int_R \int_R \cdots f_x(x)dx_{k+1} \cdots dx_d$

- $f_x(x) = \prod_{i=1}^d f_{x_i(x_i)}$ (Independent)

3.3 Multivariate Normal Distribution

Definition 1 *Standard Multivariate Normal Distribution*
$Z = (Z_1^\top, Z_2^\top, \ldots, Z_k^\top)$ *, where* $Z_1, \ldots, Z_k \sim N(0,1)$ *are independent.The density of* Z *is*

$$\begin{aligned} f(z) &= \prod_{i=1}^k f(z_i) \\ &= \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^k z_j^2\right\} \\ &= \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2} z^\top z\right\}. \end{aligned}$$

We written that $Z \sim N(0, I)$ *,* I *is the* $k \times k$ *identity matrix.*

Definition 2 *(General) Multivariate Normal Distribution*
a vector X *has a multivariate normal distribution* $X \sim N(\mu, \Sigma)$ *,it has density*

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{k/2}|(\Sigma)|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

where $|(\Sigma)|$ *denotes the determinant of* Σ *,* μ *is a vector of length* k *and* Σ *is the* $k \times k$ *symmetric,positive definite matrix.*

Lemma 1 *if* $X \sim N(0, I)$ *,* $Z = \mu + \Sigma^{1/2} X \sim N(\mu, \Sigma)$

Proof process:

$$
\begin{aligned}
\because Z &= \mu + \Sigma^{1/2} X \\
\therefore X &= g^{-1}(Z) = \Sigma^{-1/2}(Z-\mu) \\
\nabla g^{-1}(Z) &= \Sigma^{-1/2} \\
f_z(z) &= f_x(g^{-1}(z))|\mathrm{Det}(\nabla g^{-1}(Z))| \\
&= \frac{1}{(2\pi)^{k/2}} \exp\{-\frac{1}{2}(\Sigma^{-1/2}(Z-\mu))^\top \Sigma^{-1/2}(Z-\mu)\}|\mathrm{Det}(\Sigma^{-1/2})| \\
&= \frac{1}{(2\pi)^{k/2}} \exp\{-\frac{1}{2}(\Sigma^{-1/2}(Z-\mu))^\top \Sigma^{-1/2}(Z-\mu)\}|\mathrm{Det}(\Sigma)|^{-1/2} \\
&= \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(Z-\mu)^\top \Sigma^{-1}(Z-\mu)\}
\end{aligned} \tag{1}
$$

Definition 3 $\Sigma^{1/2}$ —- *the square root of* Σ
has the following properties:

- $\Sigma^{1/2}$ *is symmetric*
- $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$
- $\Sigma^{1/2}\Sigma^{-1/2} = \Sigma^{-1/2}\Sigma^{1/2} = I, where \Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$
- $\Sigma^{1/2} = UD^{1/2}U^\top$, $\Sigma^{-1/2} = UD^{-1/2}U^\top$ *where* $D^{1/2} = \begin{pmatrix} \sqrt{D_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{D_{KK}} \end{pmatrix}$

3.4 Multinomial Distribution

Considering throwing a coin which has k different faces n times.
$p = (p_1, \ldots, p_k)$, p_j : the probability of throwing a coin with face j. ($p_j \geq 0$ and $\sum_{j=1}^k p_j = 1$)
$X = (X_1, \ldots, X_k)$, X_j : the number of times that face j appears.($n = \sum_{j=1}^k X_j$)
We say that X has a $Multinomial(n,p)$ distribution written $X \sim Multinomial(n,p)$.
The probability function is

$$
\mathbb{P}(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}
$$

Lemma 2 *Suppose that* $X \sim Multinomial(n,p)$ *where* $X = (X_1, \ldots, X_k)$ *and* $p = (p_1, \ldots, p_k)$.
The marginal distribution of $X_j \sim B(n, p_j)$.

3.5 Expectation

3.5.1 Mean Value

Definition 4 *The expected value, or mean, or first moment, of X is defined to be*

$$\mathbb{E}(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Example 1 *Flip a fair coin two times. Let X be the number of heads. Then,*

$$\begin{aligned} \mathbb{E}(X) &= \int x dF_X(x) \\ &= \sum_x x f_X(x) \\ &= 0 \times f(0)) + 1 \times f(1) + 2 \times f(2) \\ &= 0 \times 1/4 + 1 \times 1/2 + 2 \times 1/4 \\ &= 1 \end{aligned} \tag{2}$$

3.5.2 The Role of the Lazy Statistician

Definition 5 *(The Role of the Lazy Statistician)Let $Y = r(X)$.Then*

$$\begin{aligned} \mathbb{E}(Y) &= \mathbb{E}(r(X)) \\ &= \int r(x) dF_X(x) \\ &= \int r(x) f_X(x) dx \end{aligned} \tag{3}$$

Example 2 *Let A be an event where $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ if $x \notin A$.Then*

$$\mathbb{E}\left(I_A(X)\right) = 0 \cdot \mathbb{P}(X \notin A) + 1 \cdot \mathbb{P}(x \in A) = \mathbb{P}(X \in A).$$