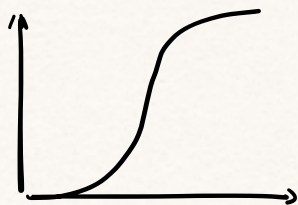
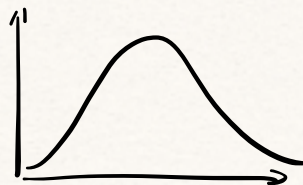


I. PDF, CDF



CDF



PDF

$$F_X(x) = P(X \leq x) \rightarrow f_X(x) = \frac{d}{dx} F(x)$$

$$F_X(x) = \int_{-\infty}^x f_X(x) dx$$

Expectation

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

$$Var[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$$

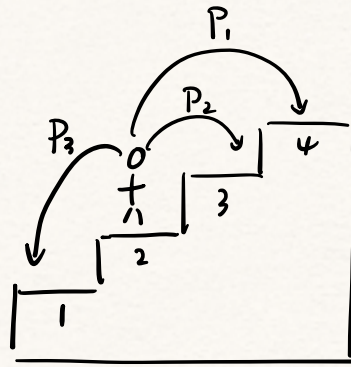
$$= E[X^2] - E[X]^2$$

Proof

$$\int_{-\infty}^{\infty} (x^2 - 2x E[X] + E[X]^2) p(x) dx$$

$$E[X^2] - 2x E[X] \underbrace{\int_{-\infty}^{\infty} x p(x) dx}_{E[X]} + E[X]^2$$

1. (a) (i)



Y : difference between ending & starting floor in 1 step

$$E[Y] = ?$$

$$Y = \begin{cases} +2 & \text{w.p. } p_1 \\ +1 & \text{w.p. } p_2 \\ -1 & \text{w.p. } p_3 \end{cases}$$

$$E[Y] = \sum_i y_i p_i = 2p_1 + p_2 - p_3$$

(a) (ii) X : final floor after 2 i.i.d. steps

$$E[X] = ?$$

$$\begin{aligned} E[X] &= 2 + E[Y] + E[Y] && \text{linearity of expectation} \\ &= 2 + 2E[Y] \end{aligned}$$

(a) (iii) X : final floor after 2 repeated steps

$$E[X] = ?$$

$$E[X] = 2 + E[2Y] = 2 + 2E[Y]$$

$$(b) \quad x_1, x_2, x_3 \in \text{Unif}[0, 1]$$

$$P\left(\frac{1}{4} \leq X_{(1)} \leq \frac{3}{4}\right) = ?$$

Fact 1. Let $X_{(j)}$ denote the j th order statistic in a sample of i.i.d. random variables; that is, the j th element when the items are sorted in increasing order $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$.

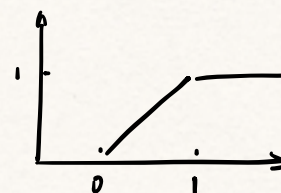
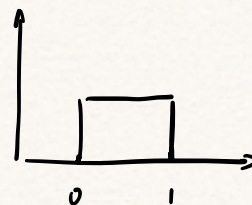
The PDF of $X_{(j)}$ is given by:

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!(j-1)!} [F(x)]^{j-1} [1-F(x)]^{n-j} f(x), \quad (1)$$

Since $X \sim \text{Unif}[0, 1]$

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x > 1 \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$



For $\frac{1}{4} \leq X_{(1)} \leq \frac{3}{4}$:

$$f_{X_{(1)}}(x) = 6x(1-x)$$

$$\begin{aligned} P\left(\frac{1}{4} \leq X_{(1)} \leq \frac{3}{4}\right) &= \int_{\frac{1}{4}}^{\frac{3}{4}} f_{X_{(1)}}(x) dx \\ &= 6 \int_{\frac{1}{4}}^{\frac{3}{4}} x(1-x) dx \\ &= 6 \int_{\frac{1}{4}}^{\frac{3}{4}} x - x^2 dx \\ &= 6 \left[\frac{1}{2} x^2 - \frac{1}{3} x^3 \right]_{\frac{1}{4}}^{\frac{3}{4}} \\ &= \frac{11}{16} \end{aligned}$$

Norms

Let $v \in \mathbb{R}^n$

$$1\text{-norm } (l_1) : \|v\|_1 = \sum_{i=1}^n |v_i|$$

$$2\text{-norm } (l_2) : \|v\|_2 = \sqrt{v^T v} = \sqrt{\sum_{i=1}^n v_i^2}$$

$$\infty\text{-norm } (l_\infty) : \|v\|_\infty = \max_i |v_i|$$

$\uparrow$ bigger = more sensitive to outliers

Matrices

Let $A \in \mathbb{R}^{n \times n}$

$$\text{Symmetric} : A = A^T$$

$$\text{quadratic form} : x^T A x \text{ for } x \in \mathbb{R}^n$$

$$\text{positive semidefinite} : x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n$$

$$2.1(a) \quad w \in \mathbb{R}^n, \quad Y \in \mathbb{R}^m \quad X = \begin{bmatrix} x_1^T \\ \vdots \\ x_m^T \end{bmatrix}$$

$$\text{show: } \|Xw - Y\|_2^2 = \sum_{i=1}^m (x_i^T w - y_i)^2$$

$$\text{By definition of } 2\text{-norm}, \|v\|_2^2 = (\sqrt{v^T v})^2 = v^T v = \sum_i v_i^2$$

$$\|Xw - Y\|_2^2 = \sum_i (x_i^T w - y_i)^2$$

(b) Let $L(w) = \|Xw - Y\|_2^2$. What is $\nabla_w L(w)$?

Gradient $\nabla_w L(w)$: direction of steepest ascent

$$\begin{bmatrix} \frac{\partial L(w)}{\partial w_1} \\ \vdots \\ \frac{\partial L(w)}{\partial w_n} \end{bmatrix}$$

Approach 1: Summation

$$\begin{aligned} \frac{\partial L(w)}{\partial w_j} &= \frac{\partial}{\partial w_j} \sum_{i=1}^m (x_i^T w - y_i)^2 \\ &= \frac{\partial}{\partial w_j} \sum_{i=1}^m \left(\left(\sum_{k=1}^n x_{ik} w_k \right) - y_i \right)^2 \\ &= \sum_{i=1}^m \frac{\partial}{\partial w_j} \left(\left(\sum_{k=1}^n x_{ik} w_k \right) - y_i \right)^2 \\ &= \sum_{i=1}^m 2 \left(\left(\sum_{k=1}^n x_{ik} w_k \right) - y_i \right) \cdot \frac{\partial}{\partial w_j} \underbrace{\sum_{k=1}^n x_{ik} w_k}_{=0 \text{ } \forall k \neq j} \end{aligned} \quad \text{chain rule}$$

$$\nabla_w L(w) = \begin{bmatrix} \sum_{i=1}^m 2 \left(\left(\sum_{k=1}^n x_{ik} w_k \right) - y_i \right) x_{i1} \\ \vdots \\ \sum_{i=1}^m 2 \left(\left(\sum_{k=1}^n x_{ik} w_k \right) - y_i \right) x_{in} \end{bmatrix}$$

Xw
 Y
 X

$$= 2 X^T (Xw - Y)$$

Matrix Calculus 101

$$y = Ax \rightarrow \frac{\partial y}{\partial x} = A$$

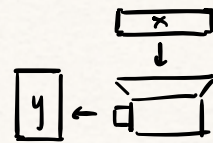
$$y = x^T A \rightarrow \frac{\partial y}{\partial x} = A^T$$

$$y = x^T x \rightarrow \frac{\partial y}{\partial x} = 2x$$

$$y = x^T A x \rightarrow \frac{\partial y}{\partial x} = (A + A^T) x = 2Ax \text{ if } A \text{ symmetric}$$

Product rule: $y = a(x)^T b(x) \rightarrow \frac{\partial y}{\partial x} = \left(\frac{\partial a(x)}{\partial x}\right)^T b(x) + \left(\frac{\partial b(x)}{\partial x}\right)^T a(x)$

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} A \end{bmatrix} \begin{bmatrix} x \end{bmatrix}$$



Approach 2: Matrix Calculus

$$\nabla_w L(w) = \nabla_w \|xw - y\|_2^2$$

$$= \nabla_w (xw - y)^T (xw - y)$$

$$= \left(\frac{\partial (xw - y)}{\partial w}\right)^T (xw - y) + \left(\frac{\partial (xw - y)}{\partial w}\right)^T (xw - y) \text{ product rule}$$

$$= x^T (xw - y) + x^T (xw - y)$$

$$= 2 x^T (xw - y)$$

2.2

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Row space : span (row vectors)

Column space : (Range(X)) : span (column vectors)

Nullspace : $\{v : Xv = 0\}$

Rank : maximum number of L.I rows / columns

2.3 (a) $\text{Row}(X) = \text{Col}(X^T)$

(b) nullspace of X and row space of X are orthogonal complements.

$$\text{Null}(X) = \text{Row}(X)^\perp$$

$$v \in \text{Null}(X) \Leftrightarrow Xv = \vec{0} \Leftrightarrow \langle x_i, v \rangle = 0 \quad \forall i$$

v perpendicular to all row vectors.

(c) $\text{Null}(X^T) = \text{Col}(X)^\perp$

2.4 (a) $X^T X$ is symmetric and PSD

(b) $\text{Null}(X) = \text{Null}(X^T X)$

Backward proof: $v \in \text{Null}(X) \Rightarrow Xv = 0 \Rightarrow X^T Xv = 0$

$$(X^T X)v = 0 \Rightarrow v^T X^T Xv = 0 \Rightarrow \|Xv\|_2^2 = 0$$

(c) $X^T X$ is not always invertible

$$\Rightarrow Xv = 0$$

(d) "tall skinny" vs "short wide" X

We'll see: argmin $\|y - Ax\|_2^2 = (X^T X)^{-1} X^T y$.

"short wide" $\rightarrow$ non empty null space $\rightarrow X^T X$ not invertible

(e) $\text{Row}(X^T X) = \text{Col}((X^T X)^T) = \text{Col}(X^T X)$

$$\text{Row}(X^T X) = \text{Null}(X^T X)^\perp = \text{Null}(X)^\perp = \text{Row}(X)$$