Notation and Basic Facts

EMALCA 2025: High-dimensional Statistics

Class of June 30

General Notation

1. For sequences of real numbers a_n and b_n we write $a_n = O(b_n)$ if $a_n \leq C b_n$ for all n and for some constant $C > 0$. An equivalent notation is sometimes used and this is expressed as $a_n \lesssim b_n$.

2. The ℓ_p for $p > 0$ norm in $\mathbb{R}^d$ is defined as $\|v\|_p = (\sum_{i=1}^d v_i^p)^{1/p}$ for $v \in \mathbb{R}^d$. When the context is clear, sometimes for the case $p = 2$ we drop the index p and simply denote the resulting norm as $\|v\|$. The ℓ_∞ norm of a vector is defined as

$$\|x\|_\infty := \max_{i=1,\ldots,d} |x_i|$$

   for $x \in \mathbb{R}^d$.

3. For a matrix $A \in \mathbb{R}^{n \times m}$ its Frobenius norm is defined as

$$\|A\|_F := \sqrt{\sum_{i=1}^n \sum_{j=1}^m A_{i,j}^2},$$

   where $A_{i,j}$ is the entry in the ith row and jth column of A.

4. Throughout, we think of vectors as column vectors, thus $x \in \mathbb{R}^n$ means

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

   and $x^\top = (x_1, \ldots, x_n)$. The inner product in $\mathbb{R}^n$ is written as

$$x^\top y = \sum_{i=1}^n x_i y_i, \quad \forall\, x, y \in \mathbb{R}^n$$

   though sometimes we denote this as $\langle x, y \rangle$.

5. For an event $\mathcal{A}_n$ depending on n, we say that $\mathcal{A}_n$ holds with high probability if

$$a_n := \mathbb{P}(\mathcal{A}_n)$$

satisfies $\lim_{n\to\infty} a_n = 1$.

6. For a collection of vectors $v_1, \ldots, v_n \subset \mathbb{R}^d$, the notation $\text{span}\{v_1, \ldots, v_n\}$ refers to the vector subspace generated by $v_1, \ldots, v_n$.

7. For a matrix $A \in \mathbb{R}^{m\times n}$ we define its kernel as

$$\text{Ker}(A) := \{v \in \mathbb{R}^n : Av = 0\}.$$

The column space of is defined as

$$\mathcal{C}(A) := \{w \in \mathbb{R}^m : w = Av \text{ for some } v \in \mathbb{R}^n\}.$$

8. For a function $f : \mathcal{C} \to \mathbb{R}$ with $\mathcal{C} \subset \mathbb{R}^n$, the notation

$$x_0 = \arg\min_{x\in\mathcal{C}} f(x)$$

indicates that $f(x_0) \leq f(x)$ for all $x \in \mathcal{C}$.

9. For a vector $\theta \in \mathbb{R}^n$ and $S \subset \{1, \ldots, n\}$, we denote the vector $\theta_S \in \mathbb{R}^{|S|}$ as the subvector of θ obtained by only keeping the entries corresponding to S. For instance, for $\theta = (1, 4, 5, 6)^\top$ and $S = \{1, 3\}$, we have that $\theta_S = (1, 5)^\top$.

10. For $n \in \mathbb{N}$, denote $[n] := \{1, \ldots, n\}$.

Random Variables

1. Let X be a continuous random variable (r.v.) $-\infty < X < \infty$.

2. The function $f(x)$ is the so called probability density function (pdf) if

   a) $f(x) \geq 0$

   b) $\int_{-\infty}^{\infty} f(x)dx = 1$,

   c) $\mathbb{P}(a < X < b) = \int_a^b f(x)dx$, $\mathbb{P}(X < b) = \int_{-\infty}^b f(x)dx$, $\mathbb{P}(a < X) = \int_a^\infty f(x)dx$.

   Note that for a continuous random variable

   $$\mathbb{P}(X \leq a) = \mathbb{P}(X < a),$$

   but this is not true for a discrete random variable.

3. Cumulative distribution function (cdf):

$$F(x) \;=\; \mathbb{P}(X \leq x).$$

Also,

$$F'(x) = f(x).$$

4. Compute probabilities using cdf:

$$\mathbb{P}(a < X < b) \;=\; \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a).$$

5. Mean of a continuous r.v.

$$\mu := E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

6. For a random varialbe X with pdf f and a function g, we have that

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

7. Variance of continuous r.v.

$$\mathrm{Var}(X) := E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2.$$

Sometimes we denote $\sigma^2 = \mathrm{Var}(X)$.

8. Some properties. Let a and b be constants and X, Y r.v. When the quantities exist, we have that:

a) $E(X+a) = a + E(X)$.

b) $E(X+Y) = E(X) + E(Y)$.

c) $\mathrm{var}(X+a) = \mathrm{var}(X)$.

d) $\mathrm{var}(aX) = a^2 \mathrm{var}(X)$.

e) If X and Y are independent then

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y).$$

9. If X is a continuous non-negative random variable, then

$$\mathbb{E}(X) \;=\; \int_{0}^{\infty} \mathbb{P}(X \geq u) du.$$

10. For a collection of events $A_1, \ldots, A_n$ the union bound inequality states that:

$$\mathbb{P}(A_1 \cup \ldots \cup A_n) \ \leq \ \sum_{i=1}^{n} \mathbb{P}(A_i).$$

11. The moment generating function of random variable X is defined as

$$M_X(t) = \mathbb{E}(\exp(tX)).$$

Informal: Therefore, If X is discrete

$$M_X(t) = \sum_x e^{tx} P(x).$$

If X is continuous

$$M_X(t) = \int_x e^{tx} f(x) dx.$$

Aside

$$e^x \ = \ 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots.$$

Similarly,

$$e^{tx} \ = \ 1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \ldots$$

Let X be a discrete random variable. Then

$$\mathbb{E}(e^{tX}) \ = \ \sum_x e^{tx} P(x) \ = \ \sum_x \left[1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \ldots\right] P(x).$$

Or

$$M_x(t) = \sum_x P(x) + \frac{t}{1!} \sum_x x P(x) + \frac{t^2}{2!} \sum_x x^2 P(x) + \ldots$$

To find the kth moment simply evaluate the kth derivative of the $M_X(t)$ at $t = 0$. Thus

$$\mathbb{E}(X^k) \ = \ M_X^{(k)}(0).$$

For example: since

$$M_x(t) = \sum_x P(x) + \frac{t}{1!} \sum_x x P(x) + \frac{t^2}{2!} \sum_x x^2 P(x) + \ldots,$$

first moment:

$$\begin{aligned} M'_X(t) \ &= \ \tfrac{d}{dt}\left(\textstyle\sum_x P(x)\right) + \tfrac{d}{dt}\left(\tfrac{t}{1!} \textstyle\sum_x x P(x)\right) + \tfrac{d}{dt}\left(\tfrac{t^2}{2!} \textstyle\sum_x x^2 P(x)\right) + \ldots \\ &= \ \sum_x x P(x) + \frac{2t}{2!} \sum_x x^2 P(x) + \ldots. \end{aligned}$$

This implies

$$M'_X(0) \ = \ \mathbb{E}(X).$$

Similarly, for the second moment

$$M''_X(t) = \sum_x x^2 P(x) + \frac{6t}{3!} \sum_x x^3 P(x) + \ldots.$$

Therefore,

$$M''_X(0) = \mathbb{E}(X^2).$$

Example: Poisson. Suppose that $X \sim \text{Poisson}(\lambda)$. Then

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} e^{tx} \mathbb{P}(X = x) \\ &= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} \\ &= \exp(-\lambda) \exp\left(e^t \lambda\right) \\ &= \exp\left(-\lambda + e^t \lambda\right). \end{aligned}$$

12. Central Limit Theorem:

Lindeberg–Lévy CLT—Suppose $X_1, X_2, X_3 \ldots$ is a sequence of i.i.d. random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$. Then, as n approaches infinity, the random variables $\sqrt{n}(\bar{X}_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$:[4]

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right).$$
