Proba y Machine Learning
2024

MW

Contents

Chapter 1

Soluciones

Ejercicio

1) Sea $X \sim N(0,1)$. Probar que para todo $a < 1$ existe $C_a < \infty$ tal que para todo $x > 0$,

$$P(X > x) \leq C_a \frac{e^{-ax^2/2}}{\sqrt{2\pi}}$$

Se puede probar mas aun, de hecho

$$P(X > x) \leq \frac{1}{x} \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

Solucion ejercicio 1:

$$P(X > x) = \int_x^{+\infty} e^{-t^2/2} dt \leq \int_x^{+\infty} (1 + t^{-2}) e^{-t^2/2} dt$$

$$\int_x^{+\infty} (1 + t^{-2}) e^{-t^2/2} dt = \int_x^{+\infty} e^{-t^2/2} dt - [t^{-1} e^{-t^2/2}]_x^{+\infty} - \int_x^{+\infty} e^{-t^2/2} dt = \frac{e^{-x^2/2}}{x}$$

Se puede ver que $\frac{e^{-x^2/2}}{x} \leq \frac{e^{-ax^2/2}}{x}, a < 1$. Entonces:

$$P(X > x) \leq \frac{e^{-ax^2/2}}{x}$$

Si $x > \sqrt{2\pi}$, esta claro que la desigualdad vale para $C_a \geq 1$. Si $x \in I = [0, \sqrt{2\pi}]$, habria que buscar C_a tal que la desigualdad valga para todos los x en este intervalo. Para hallarlo, calculamos $\arg\min_{x \in I} \frac{e^{-ax^2/2}}{\sqrt{2\pi}}$:

$$\arg\min_{x \in I} \frac{e^{-ax^2/2}}{\sqrt{2\pi}} = \sqrt{2\pi}$$

Sabiendo que $P(X > x)$ vale a lo sumo $\frac{1}{2}$, entonces:

$$\frac{1}{2} \leq C_a \frac{e^{-a(\sqrt{2\pi})^2/2}}{\sqrt{2\pi}}$$

$$C_a \geq \frac{\sqrt{2\pi}}{2} e^{-a(\sqrt{2\pi})^2/2}$$

Concluimos que con $C_a = \max\{1, \frac{\sqrt{2\pi}}{2} e^{-a(\sqrt{2\pi})^2/2}\}$ se satisface la cota pedida.

😃

Ejercicio

2) Calcular

a) $\int_{-\infty}^{+\infty} e^{-x^2/2}dx$
b) $\int_{-\infty}^{+\infty} e^{-\frac{x_1^2+...+x_n^2}{2}}dx$

Solucion ejercicio 2:

a)

$$(\int_{-\infty}^{+\infty} e^{-x^2/2}dx)^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-x^2/2}e^{-y^2/2}dxdy$$

Usamos coordenadas polares:

$$\int_0^{+\infty}\int_0^{2\pi} re^{-r^2/2}d\theta dr$$
$$2\pi\int_0^{+\infty} re^{-r^2/2}dr = 2\pi$$

Entonces:

$$(\int_{-\infty}^{+\infty} e^{-x^2/2}dx)^2 = 2\pi \implies \int_{-\infty}^{+\infty} e^{-x^2/2}dx = \sqrt{2\pi}$$

😄

Ejercicio

3) Sea $Z = (Z_1, \ldots, Z_d) \sim N(0, \mathbb{I}_d)$ y $X = (X_1, \ldots, X_d) = \frac{1}{d}(Z_1, \ldots, Z_d)$.

a) Hallar la distribucion de X.
b) Sea $\{v_1, \ldots, v_d\}$ una base ortonormal de $\mathbb{R}^d$. Probar que $E[\langle X, v_i\rangle^2] = \frac{1}{d}, \forall 1 \leq i \leq d$.
c) Probar que para todo i y para todo $\epsilon > 0$

$$P(|X \cdot v_i| > \epsilon) \leq \frac{1}{d\epsilon^2}$$

Solucion ejercicio 3:

a)

Por Teorema de Cambio de Variable:

$$f_X(x_1, \ldots, x_d) = f_Z(dx_1, \ldots, dx_d)|d^d|,$$

donde d^d es el Jacobiano de $g(x_1, \ldots, x_d) = (dx_1, \ldots, dx_d) = (z_1, \ldots, z_d)$. Entonces:

$$f_X(x_1, \ldots, x_d) = \frac{|d^d|}{\pi^{d/2}}e^{-d^2\langle x,x\rangle/2},$$

la funcion de densidad de una Normal Multivariada con $\mu = \vec{0}$ y $\Sigma = \frac{1}{d^2}\mathbb{I}$.

b)

Antes de empezar, es muy importante darse cuenta de que para la Normal Multivariada, $Cov(X_i, X_j) = 0 \implies X_i \perp X_j$ (cosa que no necesariamente es cierta para otras distribuciones). Esto se puede demostrar viendo que la convolucion se expresa como factores de cada variable X_i por separado, que no dependen de las otras X_j. Ahora si paso a la demo:

$$E[\langle v, X\rangle^2] = E[\sum_{i=1}^{d} v_i X_i \sum_{j=1}^{d} v_j X_j] = \sum_{i=1}^{d}\sum_{j=1}^{d} E[v_i v_j X_i X_j]$$

Es facil de ver que cuando $i \neq j$, $E[v_i v_j X_i X_j] = 0$ (por independencia de las variables). Por lo tanto, consideramos unicamente el caso en el que i y j son iguales:

$$\sum_{i=1}^{d} E[vi^2 X_i^2] = \sum_{i=1}^{d} vi^2 E[X_i^2] = \sum_{i=1}^{d} v_i^2 (E^2[X_i] + Var[X_i]) = \frac{1}{d^2}\sum_{i=1}^{d} v_i^2 = \frac{1}{d^2}$$

Deberia dar $\frac{1}{d}$. Preguntar.

c)

Usamos el corolario de Markov:

$$P(|X \cdot v_i| > \epsilon) \leq \frac{E[\langle X, v_i\rangle^2]}{\epsilon^2} = \frac{1}{d\epsilon^2}$$

😄

Ejercicio

4) Sean $X_1, \ldots, X_n$ iid con distribucion Normal estandar.

a) Hallar la distribucion de X_1^2
b) Hallar la distribucion de $X_1^2 + \ldots + X_n^2$

Solucion ejercicio 4:

$$P(X_1^2 < t) = P(|X_1| < \sqrt{t}) = P(-\sqrt{t} < X_1, X_1 < \sqrt{t})$$

Si llamamos φ a la funcion de densidad de X_1, entonces:

$$P(X_1^2 < t) = \int_{-\sqrt{t}}^{\sqrt{t}} \varphi(x)dx = 2\int_0^{\sqrt{t}} \varphi(x)dx$$

Denotemos a la funcion de distribucion acumulada con Φ :

$$P(X_1^2 < t) = 2\Phi(\sqrt{t}) - 1$$

Entonces:

$$f_{X_1^2}(t) = \frac{\varphi(\sqrt{t})}{\sqrt{t}} = \frac{1}{\sqrt{2\pi t}} e^{-t/2}$$

Esta distribucion es Gama $\Gamma(\frac{1}{2}, \frac{1}{2})$

😄

Ejercicio

5) Sea $f : [0,1] \to \mathbb{R}$ una funcion continua. Probar que

$$\lim_{d\to\infty} \int_0^1 \int_0^1 \cdots \int_0^1 f(\frac{x_1 + \ldots + x_d}{d}) dx_1 dx_2 \ldots dx_d = f(\frac{1}{2})$$

Solucion ejercicio 5:

Empecemos reescribiendo la expresion a calcular. Tenemos:

$$\lim_{d\to\infty} \int_{[0,1]^d} f(\frac{x_1 + \ldots + x_d}{d}) dx$$

Notar que como el volumen del hipercubo de lado $l = 1$ es 1 en dimension d, estamos calculando la esperanza de $f(g(X))$, donde $X \sim \mathcal{U}(C^d(1))$ y $g : C^d(1) \to R$, tal que $g(X) = \overline{X}$. Ademas, $X = (X_1, \ldots, X_d)$, donde $X_i \sim \mathcal{U}([0,1])$ y $X_i \perp X_j$.

$$\lim_{d\to\infty} \int_{[0,1]^d} f(\frac{x_1 + \ldots + x_d}{d}) dx = E[f(g(X))] = E[f(\overline{X})]$$

Sabiendo esto, queremos ver que

$$E[f(\overline{X})] \xrightarrow{n\to\infty} E[f(\frac{1}{2})] = f(\frac{1}{2})$$

Por LGN, sabemos que

$$f(\overline{X}) \xrightarrow{c.s} f(\frac{1}{2}) \implies f(\overline{X}) \xrightarrow{P} f(\frac{1}{2})$$

Mas aun, sabemos que $P(|f(X_i)| < c) = 1$, dado que $(X_n)_n$ son v.a. uniformes en el $[0,1]$ y f es continua en dicho intervalo. Como $[0,1]$ es compacto, esta acotado superior e inferiormente con probabilidad 1. Entonces:

$$P(|f(X_i)| < c) = 1, f(\overline{X}) \xrightarrow{P} f(\frac{1}{2}) \implies f(\overline{X}) \xrightarrow{L^1} f(\frac{1}{2}) \implies E[f(\overline{X})] \xrightarrow{n\to\infty} E[f(\frac{1}{2})]$$

😊

Ejercicio

6) Sean X e Y variables aleatorias independientes con distribucion exponencial de parametro λ. Hallar la distribucion de $Z = X + Y$ y $W = \min\{X, Y\}$

Solucion ejercicio 6:

Buscamos primero la distribucion de Z:

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = \frac{\lambda^2}{(\lambda - it)^2} = \frac{1}{(1 - \frac{it}{\lambda})^2} = \varphi_Z(t),$$

donde $Z \sim \Gamma(2, \lambda)$.

Ahora buscamos $W = \min\{X, Y\}$:

$$P(W \leq t) = 1 - P(W > t) = 1 - (1 - P(X \leq t))^2 = 1 - e^{-2\lambda t}$$

Por lo tanto, $W \sim \mathcal{E}(2\lambda)$.

😊

Ejercicio

7) Para las variables X e Y del ejercicio anterior, hallar $X|X = Y$.

Solucion ejercicio 7:

Busco distribucion conjunta de $Z = X - Y$ y X. Por Teorema de Cambio de Variable:

$$f_{X,Z}(x,z) = \lambda e^{-\lambda x} 1_{\{x>0\}} \lambda e^{-\lambda(x-z)} 1_{\{x-z>0\}}$$
$$f_{X,Z}(x,z) = \lambda^2 e^{-2\lambda x} 1_{\{x>0\}} e^{\lambda z} 1_{\{x>z\}}$$

Ahora calculo la densidad de Z:

$$f_Z(z) = e^{\lambda z} 1_{\{z>0\}} \int_z^{+\infty} \lambda^2 e^{-2\lambda x} dx + e^{\lambda z} 1_{\{z\leq 0\}} \int_0^{+\infty} \lambda^2 e^{-2\lambda x} dx$$
$$f_Z(z) = 1_{\{z>0\}} e^{\lambda z} \frac{\lambda}{2} e^{-2\lambda z} + 1_{\{z\leq 0\}} e^{\lambda z} \frac{\lambda}{2}$$

Finalmente, calculo $X|Z = 0$:

$$f_{X|Z=0}(x) = 2\lambda e^{-2\lambda x} 1_{\{x>0\}}$$

Concluyo que $X|X = Y \sim \mathcal{E}(2\lambda)$.

😄

Ejercicio

8) Una funcion suave $\phi : \mathbb{R} \to \mathbb{R}$ se dice convexa si $\phi''(x) \geq 0, \forall x \in \mathbb{R}$.

a) Probar que si ϕ es convexa, entonces

$$\phi(y) \geq \phi(x) + \phi'(x)(y - x), \forall x, y \in \mathbb{R}.$$

b) Probar que

$$\phi(\frac{x+y}{2}) \leq \frac{1}{2}\phi(x) + \frac{1}{2}\phi(y), \forall x, y \in \mathbb{R}.$$

c) Una función suave Φ:$\mathbb{R}^d \to \mathbb{R}$ se dice convexa si la matrix $(\Phi_{x_i x_j})$ es semidefinida positiva para todo $x \in \mathbb{R}^d$ $\Big($es decir, $\sum_{i,j=1}^n \Phi_{x_i x_j} \xi_i \xi_j \geqslant 0$ para todo $\xi \in \mathbb{R}^d$$\Big)$. Probar que en ese caso

$$\Phi(y) \geqslant \Phi(x) + \nabla\Phi(x) \cdot (y - x) \text{ y } \Phi\left(\frac{x+y}{2}\right) \leqslant \frac{1}{2}\Phi(x) + \frac{1}{2}\Phi(y),$$

para todo $x, y \in \mathbb{R}$

Antes de pasar a la resolucion del problema, demuestro la siguiente prop que voy a usar para el item b:

Claim 1.0.1

Sea $\phi : \mathbb{R} \to \mathbb{R}$ una funcion suave, y $x, y \in R : y > x$. Si $\frac{\phi(\frac{x+y}{2}) - \phi(x)}{\frac{y}{2} - \frac{x}{2}} > \frac{\phi(y) - \phi(x)}{y - x}$, entonces hay un intervalo $I \subset [x, y]$ donde ϕ' decrece.

Demostracion: Se puede deducir que bajo estas hipotesis

$$\frac{\phi(y) - \phi(\frac{y+x}{2})}{y - \frac{y+x}{2}} < \frac{\phi(\frac{x+y}{2}) - \phi(x)}{\frac{y}{2} - \frac{x}{2}}.$$

Ademas, por el Teorema del Valor Medio, $\exists c \in [x, \frac{x+y}{2}] / \phi'(c) = \frac{\phi(\frac{x+y}{2}) - \phi(x)}{\frac{y}{2} - \frac{x}{2}}$, y $\exists c' \in [\frac{x+y}{2}, y] / \phi'(c') = \frac{\phi(y) - \phi(\frac{x+y}{2})}{y - \frac{x+y}{2}}$. Entonces,

$$\phi'(c) > \phi'(c'), c < c',$$

lo que demuestra la prop. 😃

Solucion ejercicio 8:

a)

Consideremos el desarrollo de Taylor de orden 2 de $\phi(y)$ centrado en x:

$$\phi(y) = \phi(x) + \phi'(x)(y - x) + \frac{\phi''(\eta)}{2}(y - x)^2,$$

donde η es el error. Sabiendo que $\phi''(\eta) > 0$, entonces:

$$\phi(x) + \phi'(x)(y - x) + \frac{\phi''(\eta)}{2}(y - x)^2 \geq \phi(x) + \phi'(x)(y - x)$$

Por lo tanto:

$$\phi(y) \geq \phi(x) + \phi'(x)(y - x)$$

b)

Lo encaramos por absurdo: suponemos que $\phi(\frac{x+y}{2}) > \frac{1}{2}\phi(x) + \frac{1}{2}\phi(y)$, y usamos el resultado del item a.

$$\phi(\frac{x+y}{2}) > \frac{1}{2}\phi(x) + \frac{1}{2}\phi(y) = \phi(x) + \frac{\phi(y) - \phi(x)}{y - x}\frac{y - x}{2}$$

$$\frac{\phi(\frac{x+y}{2}) - \phi(x)}{\frac{y-x}{2}} > \frac{\phi(y) - \phi(x)}{y - x}$$

Por la prop enunciada, sabemos que $\exists c, c' : c < c', \phi'(c) > \phi'(c')$, pero ϕ' es monotona creciente dado que ϕ convexa. Por lo tanto, llegamos a un absurdo.

c)

El primer item sale de manera analoga al (a). Paso al segundo:

Sabemos que

$$\phi(y) \geq \phi(x) + \nabla\phi(x)(y - x), \forall x, y \in \mathbb{R}$$

En particular, vale que

$$\phi(x) \geq \phi(\frac{x+y}{2}) + \nabla\phi(\frac{x+y}{2})(\frac{x-y}{2}),$$
$$\phi(y) \geq \phi(\frac{x+y}{2}) + \nabla\phi(\frac{x+y}{2})(\frac{y-x}{2})$$

Sumando:

$$\phi(x) + \phi(y) \geq 2\phi(\frac{x+y}{2}) + \nabla\phi(\frac{x+y}{2})\vec{0}$$
$$\implies \phi(x) + \phi(y) \geq 2\phi(\frac{x+y}{2})$$
$$\implies \frac{\phi(x)}{2} + \frac{\phi(y)}{2} \geq \phi(\frac{x+y}{2})$$

😃

Ejercicio

10. Sea s_d el volumen de la esfera de radio uno $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d, |x| = 1\}$ de $\mathbb{R}^d$. Probar que el volumen de la esfera de radio r en $\mathbb{R}^d$ es $s_d r^{d-1}$

Solucion ejercicio 10:

El volumen de la bola unitaria en $\mathbb{R}^d$ se puede calcular con

$$B_d := \int_{-1}^{1} \dots \int_{-\sqrt{1-x_1^2-\dots-x_{d-1}^2}}^{\sqrt{1-x_1^2-\dots-x_{d-1}^2}} 1dx$$

En el caso de la bola de radio r la expresion es la misma, multiplicando los limites de integracion por r. Introduciendo el cambio de variables $rz = g(z) = x$ (con $z, x \in \mathbb{R}^d$) queda lo siguiente:

$$\int_{-r}^{r} \dots \int_{-r\sqrt{1-x_1^2-\dots-x_{d-1}^2}}^{r\sqrt{1-x_1^2-\dots-x_{d-1}^2}} 1dx = \int_{-1}^{1} \dots \int_{-\sqrt{1-z_1^2-\dots-z_{d-1}^2}}^{\sqrt{1-z_1^2-\dots-z_{d-1}^2}} J_g dz,$$

siendo J_g el jacobiano de g, que es claramente equivalente a r^d. Entonces:

$$B_d(r) = r^d B_d(1)$$

Ademas, sabemos que

$$B_d(1) = \int_0^1 \mathbb{S}^{d-1}(R)dR \implies \mathbb{S}^{d-1}(1) = \frac{\partial B_d}{\partial R}(1)$$
$$B_d(r) = \int_0^r \mathbb{S}^{d-1}(R)dR \implies \mathbb{S}^{d-1}(r) = \frac{\partial B_d}{\partial R}(r)$$

A partir de esto y la equivalencia anterior, concluimos que

$$\frac{\partial}{\partial r}B_d(r) = B_d(1)\frac{\partial}{\partial r}r^d = B_d(1)dr^{d-1}$$
$$\frac{\mathbb{S}^{d-1}(r)}{\mathbb{S}^{d-1}(1)} = \frac{B_d(1)dr^{d-1}}{B_d(1)d} = r^{d-1} \implies \mathbb{S}^{d-1}(r) = r^{d-1}\mathbb{S}^{d-1}(1)$$

😄

Ejercicio

13. Sean $X_1, \dots, X_n$ iid. Defnimos la *función de* distribución acumulada *empirica* por $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leqslant x\}$. Probar que para todo $x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_n(x) = F(x).$$

Solucion ejercicio 13:

Lo hago suponiendo que la convergencia es casi segura con respecto a ω (no puntual):

Sabemos 2 cosas: primero, que las variables $\mathbf{1}\{X_i \leq x\}$ tienen esperanza y varianza finita (dado que son Bernoulli). Ademas, son independientes, considerando que las X_n son independientes. Entonces, por LGN

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leq x\} \xrightarrow{c.s.} E[\mathbf{1}\{X \leq x\}],$$

pero $E[\mathbf{1}\{X \leq x\}] = F(x)$. O sea,

$$F_n(x) \xrightarrow{c.s.} F(x)$$

😄