Gaussian Random Variables

A specific (very useful!) continuous random variable

$X \sim \underbrace{N(\mu, \sigma^2)}$ density $\gamma_{\mu,\sigma}(x) = \dfrac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}$

Normal distribution
with mean μ, variance σ^2

$X \sim N(0,1)$ $\gamma(x) = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$

(Standard Normal Distribution)

$$\mathbb{P}[X \in [a,b]] = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx$$

What's up with $\sqrt{2\pi}$?

$$I = \int_{-\infty}^{\infty} e^{-x^2/2}$$

- I is bounded.

<u>Proof</u>: $I = 2\int_0^{\infty} e^{-x^2/2}\,dx = 2\int_0^2 \underbrace{e^{-x^2/2}}_{\leq 1}\,dx + 2\int_2^{\infty} \underbrace{e^{-x^2/2}}_{\leq e^{-x}}\,dx$

$$\leq 2\left(1 + e^{-2}\right)$$

- $I^2 = 2\pi$

<u>Proof</u>: $I^2 = \int_{-\infty}^{\infty} e^{-x^2/2}\,dx \int_{-\infty}^{\infty} e^{-y^2/2}\,dy = \iint e^{-(x^2+y^2)/2}\,dx\,dy$

$$= \iint e^{-r^2/2}\, r\,dr\,d\theta = 2\pi$$

Transformations of Gaussians

Say $X \sim N(0,1)$ $\quad Y = \sigma \cdot X + \mu$ $\quad (\sigma > 0)$

$$\mathbb{P}[Y \leq t] = \mathbb{P}\left[X \leq \frac{t-\mu}{\sigma}\right] = \int_{-\infty}^{(t-\mu)/\sigma} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx$$

$$= \int_{-\infty}^{t} \frac{e^{-(y-\mu)/2\sigma^2}}{\sqrt{2\pi}\,\sigma}\, dy$$

Density of $Y = \dfrac{e^{-(y-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma}$

Multiple independent Gaussians

$X = (x_1, \ldots, x_n)$ each $x_i \sim \mathcal{N}(0,1)$ independent

$$\gamma(x_1, \ldots, x_n) = \frac{e^{-x_1^2/2}}{\sqrt{2\pi}} \cdot \ldots \cdot \frac{e^{-x_n^2/2}}{\sqrt{2\pi}}$$

$$= \frac{e^{-(x_1^2 + \cdots + x_n^2)/2}}{(2\pi)^{n/2}}$$

$$= \frac{e^{-\|x\|^2/2}}{(2\pi)^{n/2}}$$

Linear Combinations of independent Gaussians

- X, Y independent, X has density f, Y has density g

$$P[X+Y \leq t] = \int_{-\infty}^{\infty} \int_{-\infty}^{t-x} f(x)\, g(y)\, dx\, dy$$

$z = x+y$

$$= \int_{-\infty}^{t} \Big(\underbrace{\int_{-\infty}^{\infty} f(x)\, g(z-x)\, dx}_{h(z)} \Big)\, dz$$

- $X_1, X_2 \sim N(0,1)$ independent.

$$\text{density of } X_1 + X_2 = \int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \frac{e^{-(z-x)^2/2}}{\sqrt{2\pi}}\, dx$$

$$= \Big(\int_{-\infty}^{\infty} \underbrace{e^{-x^2+zx}}_{e^{-(x-\frac{z}{2})^2 + z^2/4}}\, dx \Big) \cdot \frac{e^{-z^2/2}}{2\pi} = \frac{e^{-z^2/4}}{\sqrt{2}\cdot\sqrt{2\pi}}$$

Ex: For $X_1, \ldots, X_n$ i.i.d. $N(0,1)$, show that $\sum c_i \cdot X_i \sim N(0, \|c\|^2)$

Gaussian Expectations

Let $X \sim \mathcal{N}(0,1)$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot \gamma(x)\, dx = \int_{-\infty}^{\infty} x \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}} = 0 = \mathbb{E}[X^{2k+1}]$$

for any $k \in \mathbb{N}$

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx = 1$$

$$\mathbb{E}[e^{\lambda X^2}] = \int_{-\infty}^{\infty} e^{\lambda x^2} \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx = \int_{-\infty}^{\infty} \frac{e^{-(1-2\lambda)x^2/2}}{\sqrt{2\pi}}\, dx \quad \Big\} y = \sqrt{1-2\lambda}\, x$$

$(\lambda < \frac{1}{2})$

$$= \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \frac{dy}{\sqrt{1-2\lambda}} = \frac{1}{\sqrt{1-2\lambda}}$$

Johnson - Lindenstrauss dimension reduction [JL 84]

- Let $P \subseteq \mathbb{R}^d$ be a set of n points. $\forall \epsilon > 0$

  $\exists$ a linear transformation $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ s.t.

  $\forall\, u, v \in P$ $\quad (1-\epsilon)\|u-v\|^2 \leq \|\varphi(u) - \varphi(v)\|^2 \leq (1+\epsilon)\cdot\|u-v\|^2$

  and $k \leq \dfrac{12 \ln n}{\epsilon^2 - \epsilon^3}$

Random φ :

$$\varphi = \frac{1}{\sqrt{k}} \overbrace{\begin{bmatrix} & & \\ & \quad G \quad & \\ & & \end{bmatrix}}^{d} \left.\vphantom{\begin{bmatrix} a \\ a \\ a \end{bmatrix}}\right\} k$$

$\forall i, j \quad G_{ij} \sim \mathcal{N}(0,1)$ i.i.d

Studying just one pair (u,v)

Suffices to show that for any fixed $u,v \in P$

$$\mathbb{P}\left[(1-\epsilon)\|u-v\|^2 \le \|\varphi(u)-\varphi(v)\|^2 \le (1+\epsilon)\cdot\|u-v\|^2\right] \geq 1-\frac{2}{n^3}$$

$\varphi = \frac{1}{\sqrt{k}} \cdot G$

Say we can prove above. Then,

$$\mathbb{P}\left[\forall u,v \in P \;\; (1-\epsilon)\|u-v\|^2 \le \|\varphi(u)-\varphi(v)\|^2 \le (1+\epsilon)\cdot\|u-v\|^2\right].$$

$$\geq 1-\binom{n}{2}\cdot\frac{2}{n^3}$$

$$\geq 1-\frac{1}{n}$$

Studying a single unit vector.

Let $w = \dfrac{u-v}{\|u-v\|}$

$$\varphi = \frac{1}{\sqrt{k}} G$$

$$(1-\epsilon)\|u-v\|^2 \le \|\varphi(u)-\varphi(v)\|^2 \le (1+\epsilon)\cdot\|u-v\|^2$$

$$\Updownarrow$$

$$(1-\epsilon)\cdot\cancel{\|u-v\|^2} \le \left\| \frac{1}{\sqrt{k}} G \frac{(u-v)}{\|u-v\|} \right\|^2 \le (1+\epsilon)\cdot\cancel{\|u-v\|^2}$$

$$\Updownarrow$$

$$k(1-\epsilon) \le \|Gw\|^2 \le k(1+\epsilon)$$

Suffices to prove $\quad \underset{G}{P}\left[(1-\epsilon)k \le \|Gw\|^2 \le (1+\epsilon)k\right] \ge 1 - \dfrac{2}{n^3}$

Understanding a single w

Fix $w \in \mathbb{R}^d$, $\|w\| = 1$

$$\begin{bmatrix} & & \\ & G & \\ & & \end{bmatrix} \begin{bmatrix} \\ w \\ \\ \\ \end{bmatrix} = \underset{x}{\begin{bmatrix} \\ \\ \\ \end{bmatrix}}$$

$$x_i = G_{i1} \cdot w_1 + \cdots + G_{id} \cdot w_d$$

$$x_i \sim \mathcal{N}(0, \|w\|^2)$$

$$\sim \mathcal{N}(0, 1)$$

$$Z = \|Gw\|^2 = x_1^2 + \cdots + x_k^2$$

Sums of squares of independent Gaussians

$$Z = X_1^2 + \cdots + X_k^2 \qquad \underbrace{X_1, \ldots, X_k}_{\text{i.i.d}} \sim \mathcal{N}(0,1)$$

$$\mathbb{E}[Z] = \mathbb{E}[X_1^2] + \cdots + \mathbb{E}[X_k^2] = k$$

$$P[Z \geq (1+\epsilon)\cdot k] = \mathbb{P}[e^{\lambda Z} \geq e^{(1+\epsilon)k}]$$

$$\leq \frac{\mathbb{E}[e^{\lambda Z}]}{e^{(1+\epsilon)\cdot k}} = \prod_{i=1}^{k} \frac{\mathbb{E}[e^{\lambda X_i^2}]}{e^{(1+\epsilon)\cdot k}}$$

$$\mathbb{E}[e^{\lambda X_i^2}] = \frac{1}{\sqrt{1-2\lambda}}$$

$(\lambda < 1/2)$

Choosing λ and k

$$\mathbb{P}[Z \geq (1+\epsilon)k] \leq \frac{\left(1/\sqrt{1-2\lambda}\right)^k}{e^{\lambda(1+\epsilon)k}} = \left(\frac{e^{-2\lambda(1+\epsilon)}}{1-2\lambda}\right)^{k/2} \leq \frac{1}{n^3}$$

using $\lambda = \dfrac{\epsilon}{2(1+\epsilon)}$

$$\mathbb{P}[Z \geq (1+\epsilon)k] \leq \left((1+\epsilon)\cdot e^{-\epsilon}\right)^{k/2}$$

$$\leq (1+\epsilon)\left(1-\epsilon+\frac{\epsilon^2}{2}\right)$$

$$= 1 - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{2} \leq e^{-(\epsilon^2-\epsilon^3)/2}$$

$\mathbb{P}[\,..\,] \leq \dfrac{1}{n^3}$ for $k = \dfrac{12 \ln n}{\epsilon^2 - \epsilon^3}$

Ex: Obtain similar bound for $\mathbb{P}[Z \leq (1-\epsilon)k]$