Lecture 3

1D gaussian

$$\underset{\substack{\uparrow\\ \text{pdf}}}{\mathcal{N}}(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

$$\underset{\substack{\uparrow\\ \text{cdf}}}{\Phi}(y|\mu,\sigma^2) = \int_{-\infty}^{y} dz\, \mathcal{N}(z|\mu,\sigma^2) =$$

$$= \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right],$$ where

$z = \dfrac{y-\mu}{\sigma}$ and $\operatorname{erf}(u) = \dfrac{2}{\sqrt{\pi}}\displaystyle\int_0^u dt\, e^{-t^2}$ is the error function.

If $Y = \mathcal{N}(y|\mu,\sigma^2)$,

$$\begin{cases} E[Y] = \mu, \\ E[Y^2] = \mu^2 + \sigma^2, \\ \operatorname{std}[Y] = \sigma. \end{cases}$$

Regression: $p(y|\vec{x},\vec{\theta}) = \mathcal{N}(y|f_\mu(\vec{x},\vec{\theta}), f_\sigma^2(\vec{x},\vec{\theta}))$,

where $f_\mu \in \mathbb{R}$, $f_\sigma \in \underbrace{\mathbb{R}_+}_{\text{non-negative}}$

For ex.,
$$\begin{cases} f_\mu(\vec{x},\vec{\theta}) = \vec{w}_\mu^T \cdot \vec{x} + b, \\ f_\sigma(\vec{x},\vec{\theta}') = \log(1 + e^{\vec{w}_\sigma^T \cdot \vec{x}}). \end{cases}$$

$\sigma_+(a) = \log(1+e^a)$ is the softplus function.

Central limit theorem

Consider N iid rv's: $X_n \sim p(X)$, $n=1,\dots,N$

where $\begin{cases} E[X] = \mu, \\ Var[X] = \sigma^2. \end{cases}$

Next, define $S_N = \sum_{n=1}^{N} X_n$

as $N \uparrow$, $p(S_N)$ is closely approximated by

$$p(S_N = u) = \mathcal{N}(u \mid \mu N, \sigma^2 N).$$

In other words, $Z_N = \dfrac{S_N - N\mu}{\sigma\sqrt{N}} = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{N}}$,

where $\bar{X} = \dfrac{S_N}{N}$ is the sample mean,

is distributed as $\mathcal{N}(Z_N \mid 0, 1)$.

RV transforms

Consider $\begin{cases} y = f(x), \\ x = f^{-1}(y). \end{cases}$

Then $\overset{\text{cdf}}{P_y}(y) = P(\underbrace{f(X)}_{Y} \leq y) = P(X \leq f^{-1}(y)) =$

$$= P_x(\underbrace{f^{-1}(y)}_{x}).$$

Now $\underset{\text{pdf}}{p(y)} = \dfrac{d}{dy} P_y(y) = \dfrac{d}{dy} P_x(x) = \pm \dfrac{dx}{dy} \overbrace{\dfrac{d}{dx} P_x(x)}^{p(x)} =$

to keep $p(y)$ pos.

$$= \pm\left(\frac{d}{dy} f^{-1}(y)\right) p(x).$$

Finally, $\quad p(y) = p(f^{-1}(y)) \left| \frac{d}{dy} f^{-1}(y) \right|$.

Multi-D extension:

$$\begin{cases} \vec{y} = \vec{f}(\vec{x}), & \mathbb{R}^n \to \mathbb{R}^n \\ \vec{x} = \vec{g}(\vec{y}) \end{cases}$$

Then $\quad p(\vec{y}) = p(\vec{g}(\vec{y}))\ |\det J(\vec{y})|$, where

$$J(\vec{y}) = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \cdots & & \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}.$$

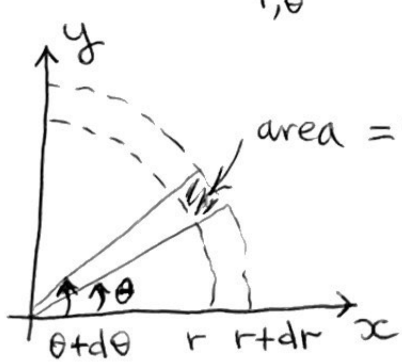
<u>Ex.</u> (2D) $\quad (x_1, x_2) \to (\overset{y_1}{r}, \overset{y_2}{\theta})$

$$\begin{cases} x_1 = r\cos\theta, \\ x_2 = r\sin\theta. \end{cases}$$

$$J = \begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},$$

$\det J = r$.

Hence $\quad p_{r,\theta}(r,\theta) = p_{x_1,x_2}(r\cos\theta, r\sin\theta)\ \underline{\underline{r}}$.



area $= r\,dr\,d\theta \Rightarrow p_{r,\theta}(r,\theta)\,dr\,d\theta =$

$= p_{x,y}(\underset{r\cos\theta}{\underset{\shortparallel}{x}}, \underset{r\sin\theta}{\underset{\shortparallel}{y}}) \times r\,dr\,d\theta$.

Moments

Consider

$$y = f(\vec{x}) = \vec{a}^T \cdot \vec{x} + b, \quad \text{then}$$

$$E[\vec{x}] = \vec{\mu} \Rightarrow E[y] = \vec{a}^T \cdot \vec{\mu} + b$$

$$\underset{\vec{x} \in R^D}{\mathrm{Cov}[\vec{x}]} = \Sigma_{D\times D} \Rightarrow \mathrm{Var}[y] = \vec{a}^T \Sigma_{D\times D} \vec{a}.$$

$$\mathrm{Cov}[\vec{x}] = E[(\vec{x} - E[\vec{x}]) \cdot (\vec{x} - E[\vec{x}])^T].$$

Thus, $\Sigma_{ij} = E[(x_i - E[x_i])(x_j - E[x_j])]$.

Ex. $y = f(x_1, x_2) = x_1 + x_2 \Rightarrow \vec{a} = (1,1), \quad b = 0$

Then

$$\mathrm{Var}[y] = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \underset{=\Sigma_{12}}{\Sigma_{21}} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} =$$

$$= \Sigma_{11} + \Sigma_{22} + 2\Sigma_{12} = V[x_1] + V[x_2] + 2\,\mathrm{Cov}[x_1, x_2]$$

Convolution theorem

Consider indep. rv's x_1 & x_2: $\begin{cases} p_1(x_1), \\ p_2(x_2). \end{cases}$

What is $p(y)$, where $y = x_1 + x_2$?

$$\underset{\text{cdf}}{P(y < \tilde{y})} = \int_{-\infty}^{\infty} dx_1\, p_1(x_1) \int_{-\infty}^{\tilde{y} - x_1} dx_2\, p_2(x_2).$$

$$x_1 + x_2 < \tilde{y} \Rightarrow x_2 < \tilde{y} - x_1$$

Then $$p(\tilde{y}) = \frac{d P(y<\tilde{y})}{d\tilde{y}} = \int_{-\infty}^{\infty} dx_1 \, p_1(x_1) \frac{d}{d\tilde{y}}\left[\int_{-\infty}^{\tilde{y}-x_1} dx_2 \, p_2(x_2)\right]$$ (11)

$$= \int_{-\infty}^{\infty} dx_1 \, p_1(x_1) \, p_2(\tilde{y}-x_1).$$

Thus, $$\boxed{p(y) = \int_{-\infty}^{\infty} dx \, p_1(x) \, p_2(y-x)}.$$ convolution

Ex. $$\begin{cases} x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \end{cases} \Rightarrow p(\underbrace{x_1+x_2}_{y}) = \mathcal{N}(y \mid \mu_1+\mu_2, \sigma_1^2+\sigma_2^2).$$

$\mathrm{Cov}[x_1, x_2] = 0$, indep. variables