Statistical Data Analysis
Discussion notes – week 2

- Transformation of variables with delta function
- Examples of transformation of variables
- Example of error propagation
- Mean and variance of Poisson mixture model
- Proof the covariance matrix is positive semi-definite

Example of transformation of variables with Mellin convolution

Consider the joint pdf

$$f(x,y) = 1 \quad , \quad 0 \le x \le 1 \ , \ 0 \le y \le 1$$

a) Find pdf of $z = xy$

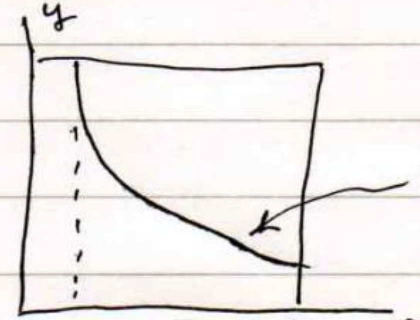
In lectures we found

$$g(z) = \int f\left(x, \frac{z}{x}\right) \frac{dx}{x}$$ (Mellin convolution)

$$= \int_{x_{min}}^{x_{max}} 1 \cdot \frac{dx}{x}$$

$f(x,y)$ is nonzero for $0 \le x \le 1,\ 0 \le y \le 1$

$$\Rightarrow\ 0 \le \frac{z}{x} \le 1 \quad \Rightarrow \quad 0 \le z \le x$$



$$\Rightarrow\ x_{min} = z$$

$$x_{max} = 1$$

$$\Rightarrow\ g(z) = \int_z^1 \frac{dx}{x} = \ln x \Big|_z^1$$

$$= -\ln z\ , \quad 0 < z \le 1$$

b) Alternative method – let

$$z = xy \qquad \Rightarrow \qquad x = u$$

$$u = x \qquad\qquad\qquad y = \frac{z}{u}$$

Jacobian is

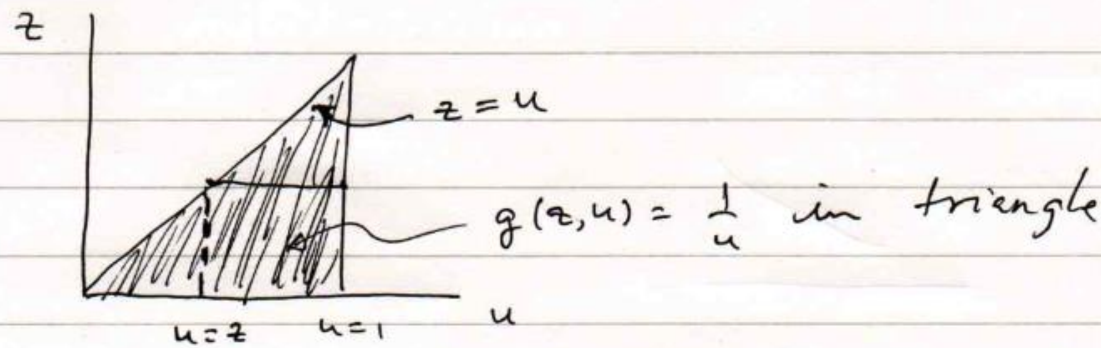
$$J = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial u} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ \frac{1}{u} & -\frac{z}{u^2} \end{vmatrix} = -\frac{1}{u}$$

$$g(z,u) = |J|\, f(x(z,u), y(z,u))$$

$$= \frac{1}{u}, \quad 0 \leq u \leq 1, \quad 0 \leq z \leq u$$

Because: $0 \leq x \leq 1 \Rightarrow 0 \leq u \leq 1$

$0 \leq y \leq 1 \Rightarrow 0 \leq \frac{z}{u} \leq 1 \Rightarrow 0 \leq z \leq u$



$$g_z(z) = \int g(z,u)\, du = \int_z^1 \frac{du}{u} = -\ln z, \quad 0 < z \leq 1$$

$\uparrow u \geq z$ (see above)

Variable transformation w/ δ function

cf. D. Gillespie, Am. J. Phys. 51 (1983) 520,

Harrison Prosper INFN School of Statistics

lectures on probability (2024)

Theorem : $\vec{x} \sim f(\vec{x})$ $\vec{x} = (x_1, \dots, x_n)$

proof

$\vec{y} = \vec{a}(\vec{x})$ $\vec{y} = (y_1, \dots, y_m)$

pdf $\longrightarrow$ $$\boxed{g(\vec{y}) = \int d^n\vec{x}\; f(\vec{x})\; \delta^{(m)}(\vec{y} - \vec{a}(\vec{x}))}$$

Example: $f(x,y)$ $\qquad z = xy$

$$g(z) = \iint dx\, dy\, f(x,y)\, \delta(\underbrace{z - xy}_{\hookrightarrow\, h(y) \equiv z - xy})$$

Use $\delta(h(y)) = \dfrac{\delta(y - y_0)}{|h'(y_0)|}$, y_0 = zero of h

$$h(y_0) = z - xy_0 = 0 \Rightarrow y_0 = \frac{z}{x}, \quad h'(y_0) = -x$$

$$g(z) = \int dx \int dy\, f(x,y) \frac{\delta(y - \frac{z}{x})}{|-x|}$$

$$= \int f\left(x, \frac{z}{x}\right) \frac{dx}{x}$$

same as before

Example 2 – error propagation

Consider r.v.s x_i, $i = 1, 2$

with $\mu_i = 10$, $\sigma_i = 1$, $\mathrm{cov}[x_i, x_j] = 0$

and let $y = \frac{x_1^2}{x_2}$. Find variance $V[y]$.

$$V[y] \approx \left(\frac{\partial y}{\partial x_1}\right)^2\bigg|_{\vec{x}=\vec{\mu}} \sigma_1^2 + \left(\frac{\partial y}{\partial x_2}\right)^2\bigg|_{\vec{x}=\vec{\mu}} \sigma_2^2$$

$$= \left(\frac{2x_1}{x_2}\right)^2\bigg|_{\vec{x}=\vec{\mu}} \sigma_1^2 + \left(-\frac{x_1^2}{x_2^2}\right)^2\bigg|_{\vec{x}=\vec{\mu}} \sigma_2^2$$

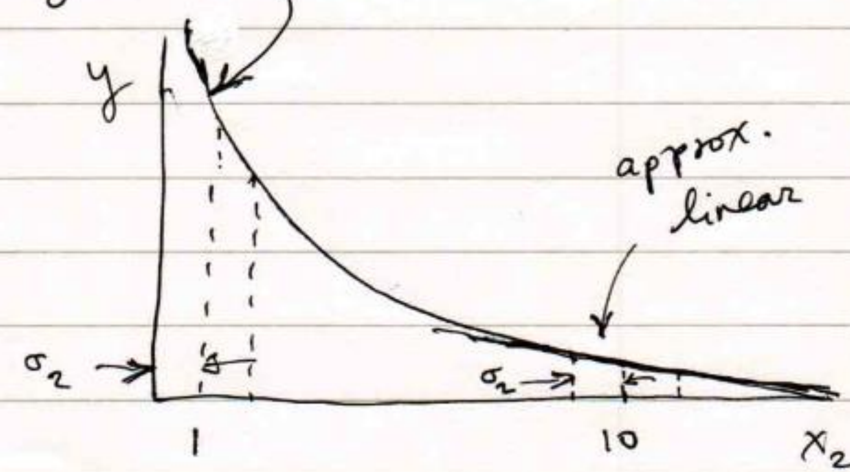
$$= \frac{4\mu_1^2}{\mu_2^2}\sigma_1^2 + \frac{\mu_1^4}{\mu_2^4}\sigma_2^2$$

$$= 4 \times 1 + 1 \cdot 1 = \underline{5} \quad \Rightarrow \quad \sigma_y = \sqrt{5} \approx 2.2$$

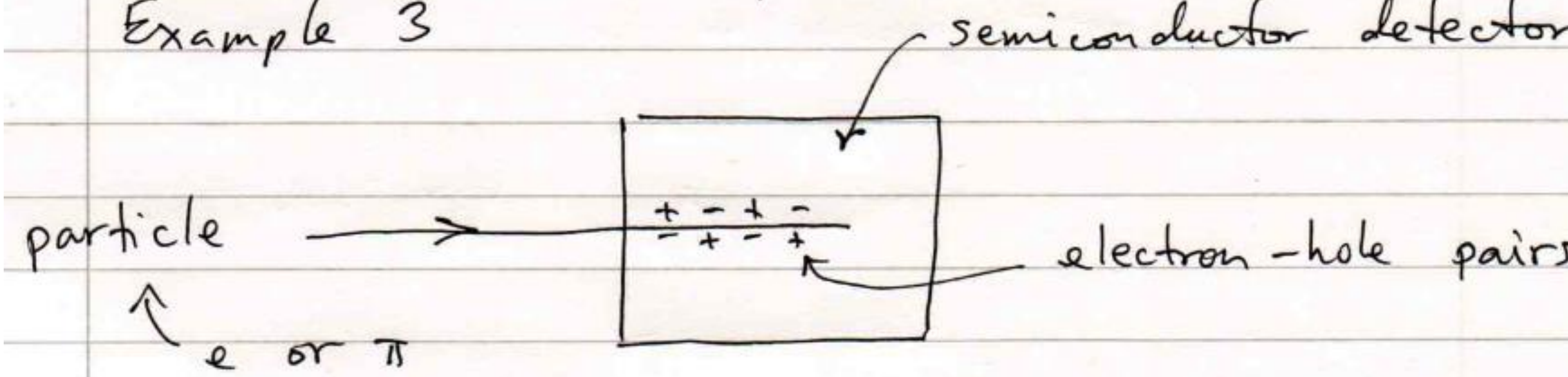
Now suppose $\mu_1 = 10$, $\mu_2 = 1$ $\quad (\sigma_1 = \sigma_2 = 1)$

Then $y = \frac{x_1^2}{x_2}$ is significantly nonlinear in a region of $\sim \pm \sigma_2$ & therefore linear err. prop. is poor approx.

$(\sigma_y \to \sqrt{10400} = 102.0)$



Example 3



$$\# \text{ of } e^-/\text{hole pairs } n \quad \begin{cases} \sim \text{Poisson}(\nu_e) & \text{for } e \\ \sim \text{Poisson}(\nu_\pi) & \text{for } \pi \end{cases}$$

Incident particles are mixture of e, π

w/ relative fractions π_e, $\underbrace{\pi_\pi}_{=1-\pi_e}$ (prior probs.)

From law of total probability

$$P(n) = P(n|\nu_\pi)\,\pi_\pi + P(n|\nu_e)\,\pi_e$$

$$= \frac{\nu_\pi^n}{n!} e^{-\nu_\pi} \pi_\pi + \frac{\nu_e^n}{n!} e^{-\nu_e} \pi_e$$

Expectation value of n is

$$E[n] = \sum_{n=0}^{\infty} n\, P(n)$$

$$= \pi_\pi \underbrace{\sum_{n=0}^{\infty} n\, P(n|\nu_\pi)}_{= E[n|\nu_\pi] = \nu_\pi} + \pi_e \underbrace{\sum_{n=0}^{\infty} n\, P(n|\nu_e)}_{= E[n|\nu_e] = \nu_e}$$

$$= \pi_\pi \nu_\pi + \pi_e \nu_e$$

Find variance $V[n] = E[n^2] - (E[n])^2$

First find

$$E[n^2] = \sum_{n=0}^{\infty} n^2 \left(P(n|\nu_\pi)\pi_\pi + P(n|\nu_e)\pi_e \right)$$

$$= \pi_\pi E[n^2|\nu_\pi] + \pi_e E[n^2|\nu_e]$$

Use fact that

$$E[n^2] = V[n] + (E[n])^2$$

and for Poisson var. $V[n] = E[n]$

$$\Rightarrow E[n^2|\nu_i] = \nu_i + \nu_i^2 \quad , \qquad i = \pi, e$$

Use fact that

$$E[n^2] = V[n] + (E[n])^2$$

and for Poisson var. $V[n] = E[n]$

$$\Rightarrow \quad E[n^2 | \nu_i] = \nu_i + \nu_i^2 \,, \qquad i = \pi, e$$

Assembling the ingredients,

$$V[n] = \pi_\pi (\nu_\pi + \nu_\pi^2) + \pi_e (\nu_e + \nu_e^2)$$

$$- (\pi_\pi \nu_\pi + \pi_e \nu_e)^2$$

Example 4 – proof that covariance matrix

$V_{ij} = \mathrm{cov}[x_i, x_j]$ is positive semi-definite

i.e. $\vec{z}^T V \vec{z} \geq 0$ for any $\vec{z} \in \mathbb{R}^n$

Can transform r.v.s to have mean of zero

i.e. let $x_i \to x_i - \mu_i$ so that

$$V = E[\vec{x}\,\vec{x}^T] \quad , \quad \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

"outer product"

Let $\vec{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{R}^n$ (const.)

$$\vec{z}^T V \vec{z} = \vec{z}^T E[\vec{x}\,\vec{x}^T]\,\vec{z}$$

$$= E[\vec{z}^T \vec{x}\,\vec{x}^T \vec{z}] \qquad \text{since } E[\;] \text{ linear}$$

$$= E[(\vec{x}^T z)^T (\vec{x}^T \vec{z})] \qquad \text{since } A^T B = ((A^T B)^T)^T$$

$$= (B^T A)^T$$

$$= E[(\vec{z}^T \vec{x})^2] \geq 0$$

real scalar

Q.E.D.

For e.g. $z_i = \delta_{ij} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ ← position j

$$\Rightarrow E[x_j^2] = V[x_j] \geq 0 .$$