# Statistical Data Analysis
# Discussion notes – week 4

- Problem sheet 1
- Test of Gaussian mean
- Chain rule for pdfs

# Problem sheet 1

**Exercise 1:** A beam of particles consists of a fraction $10^{-4}$ electrons and the rest photons. The particles pass through a double-layered detector which gives signals in either zero, one or both layers. The probabilities of these outcomes for electrons (e) and photons ($\gamma$) are

$$
\begin{aligned}
P(0\,|\,\mathrm{e}) &= 0.001 & \quad\text{and}\quad && P(0\,|\,\gamma) &= 0.99899 \\
P(1\,|\,\mathrm{e}) &= 0.01 & && P(1\,|\,\gamma) &= 0.001 \\
P(2\,|\,\mathrm{e}) &= 0.989 & && P(2\,|\,\gamma) &= 10^{-5}\,.
\end{aligned}
$$

(a) [4 marks] What is the probability for the particle to be a photon given a detected signal in one layer only?

1) Use Bayes' thm

$$P(A|B) = \frac{P(B|A)\,P(A)}{\sum_i P(B|A_i)\,P(A_i)}$$

a) $$P(\gamma|1) = \frac{P(1|\gamma)\,P(\gamma)}{P(1|\gamma)\,P(\gamma) + P(1|e)\,P(e)}$$

$$= \frac{0.001 \times 0.9999}{0.001 \times 0.9999 + 0.01 \times 10^{-4}} = 0.999\;(000899)$$

$1 - 10^{-4}$

# 1(b)

(b) [4 marks] What is the probability for a particle to be an electron given a detected signal in both layers?

b) $$P(e|2) = \frac{P(2|e)\,P(e)}{P(2|e)\,P(e) + P(2|\gamma)\,P(\gamma)}$$

$$= \frac{0.989 \times 0.0001}{0.989 \times 0.0001 + 10^{-5} \cdot (1 - 0.0001)}$$

$$= \underline{0.90818}$$

**Exercise 2:** Consider the joint probability density for two continuous variables $x$ and $y$ given by

$$f(x,y) = \begin{cases} x+y & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) [6 marks] Find the marginal pdfs $f_x(x)$ and $f_y(y)$ and indicate what they look like with a simple sketch. Are $x$ and $y$ independent? Explain.

a) $f_x(x) = \int f(x,y)\, dy$

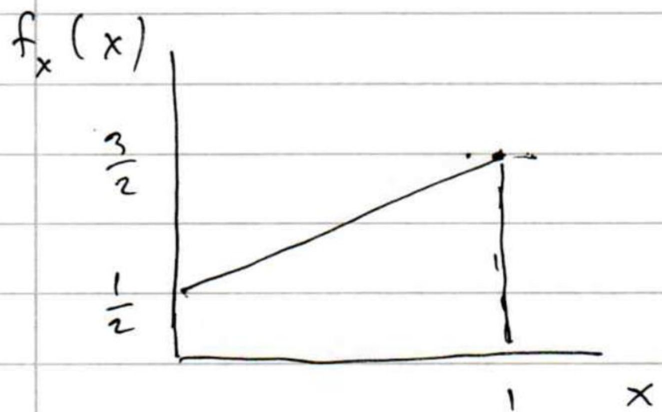
$= \int_0^1 (x+y)\, dy$

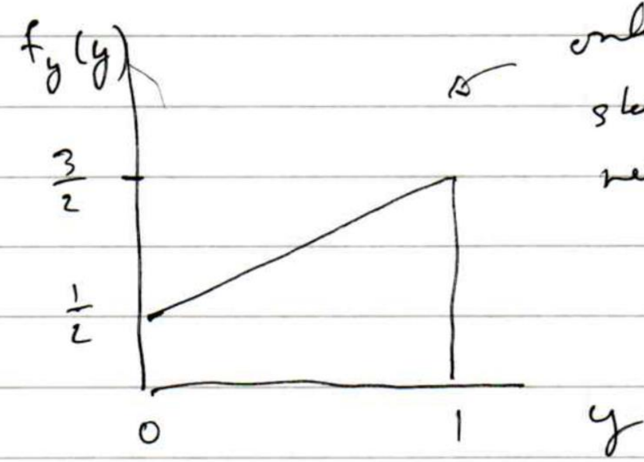
$= \left( xy + \frac{y^2}{2} \right) \Big|_0^1 = \frac{1}{2} + x,$

$0 \le x \le 1$

# 2(a) cont.

By symmetry, $f_y(y) = \frac{1}{2} + y, \quad 0 \le y \le 1$




only 1 sketch needed.

$$f(x,y) = x + y$$

$$\neq f_x(x) f_y(y) = \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right)$$

$\Rightarrow$ $x, y$ not independent.

# 2(b)

(b) [6 marks] Find the conditional probabilities $f(x|y)$ and $f(y|x)$. State how these two densities are related by Bayes theorem, and demonstrate that the relation holds using the conditional pdfs you have found together with the marginal pdfs from (a).

2b) $f(x|y) = \dfrac{f(x,y)}{f_y(y)}$

$$= \frac{x+y}{\frac{1}{2}+y}, \quad 0 \leq x \leq 1$$

By symmetry

$$f(y|x) = \frac{x+y}{\frac{1}{2}+x}, \quad 0 \leq y \leq 1$$

# 2(b) cont.

Bayes' thm:

$$f(x|y) = \frac{f(y|x)\, f_x(x)}{f_y(y)}$$

$$\frac{x+y}{\frac{1}{2}+y} \stackrel{?}{=} \frac{\frac{x+y}{\cancel{\frac{1}{2}+x}} \cdot \cancel{\left(\frac{1}{2}+x\right)}}{\frac{1}{2}+y} \quad \checkmark$$

From Week 4 extra slides

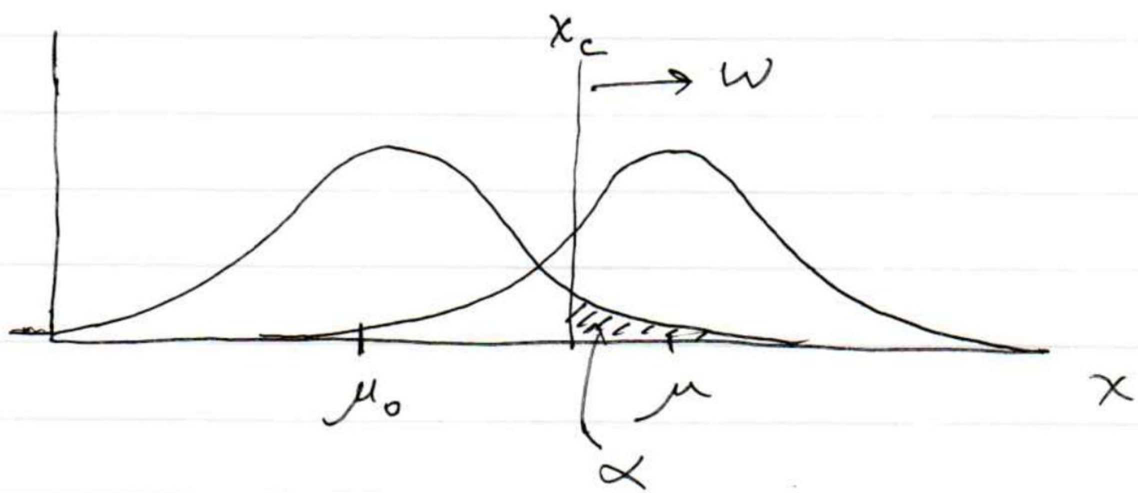
want to test

$x \sim \text{Gauss}(\mu, \sigma)$

known.

$H_0 : \mu = \mu_0$ . vs. $H_1 : \mu > \mu_0$

Take $w = \{ x : x \geq x_c \}$



$$\alpha = P(x \geq x_c \mid \mu_0)$$

$$= \int_{x_c}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu_0)^2/2\sigma^2} dx$$

let $y = \dfrac{x-\mu_0}{\sigma}$

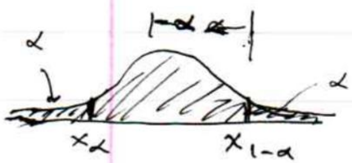
$$= \int_{\frac{x_c-\mu_0}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}\,\cancel{\sigma}} e^{-y^2/2} \cdot \cancel{\sigma}\, dy$$

$$= 1 - \Phi\left(\frac{x_c-\mu_0}{\sigma}\right)$$

Standard Gauss. cumul. dist.

$$\Rightarrow \quad x_c = \mu_0 + \sigma\,\Phi^{-1}(1-\alpha)$$

Std. Gauss. quantile



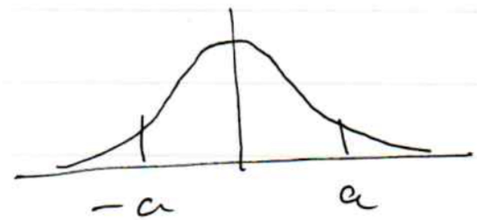

$$\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$$

Power $M = P(x \geq x_c \,|\, \mu)$

$$= \int_{x_c}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

$$= 1 - \Phi\left(\frac{x_c - \mu}{\sigma}\right)$$

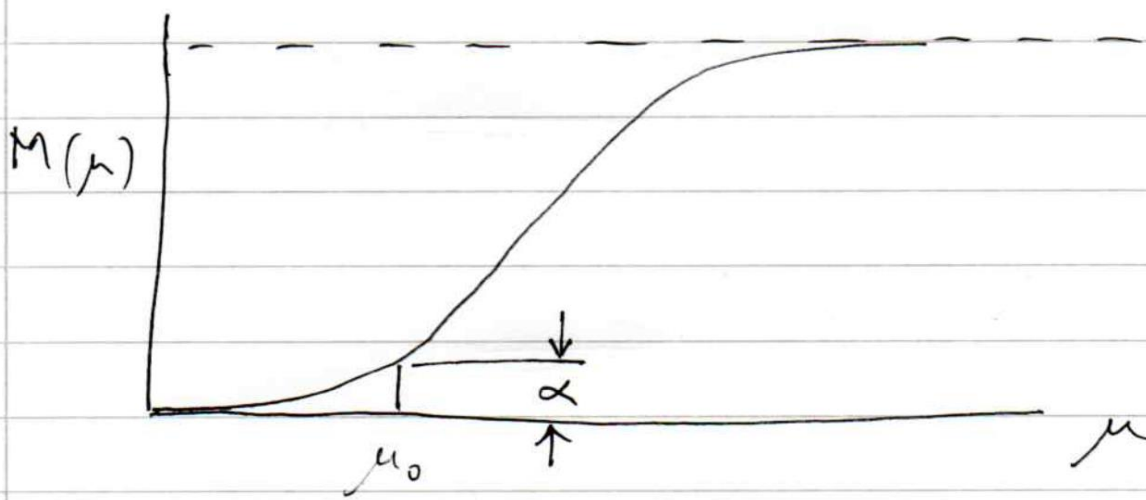
$$= \Phi\left(\frac{\mu - x_c}{\sigma}\right)$$




$$\Phi(-a) = 1 - \Phi(a)$$

$$= \Phi\left(\frac{\mu - \mu_0 - \sigma\Phi^{-1}(1-\alpha)}{\sigma}\right)$$

$$= \Phi\left(\frac{\mu - \mu_0}{\sigma} + \Phi^{-1}(\alpha)\right)$$



"Two-sided" test:

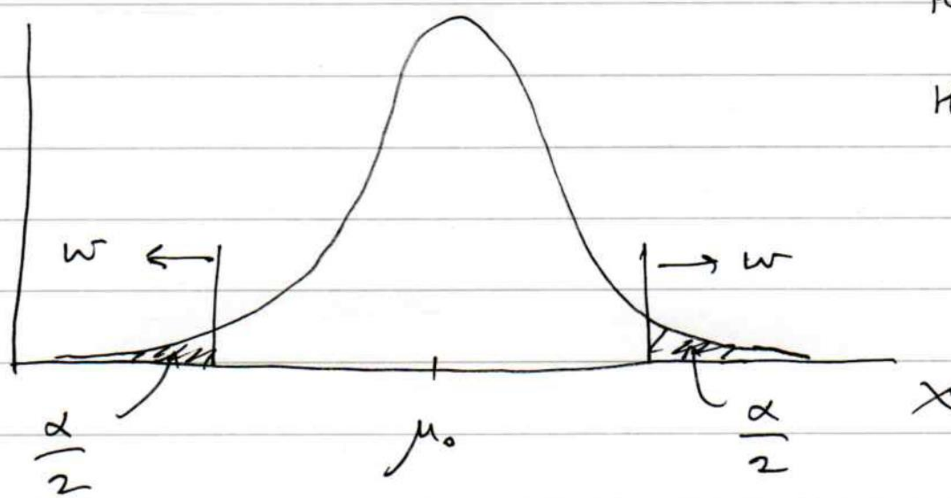
We might consider $\mu < \mu_0$ <u>and</u> $\mu > \mu_0$

as equally relevant alternatives.

$\Rightarrow$ choose two-sided critical region:

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

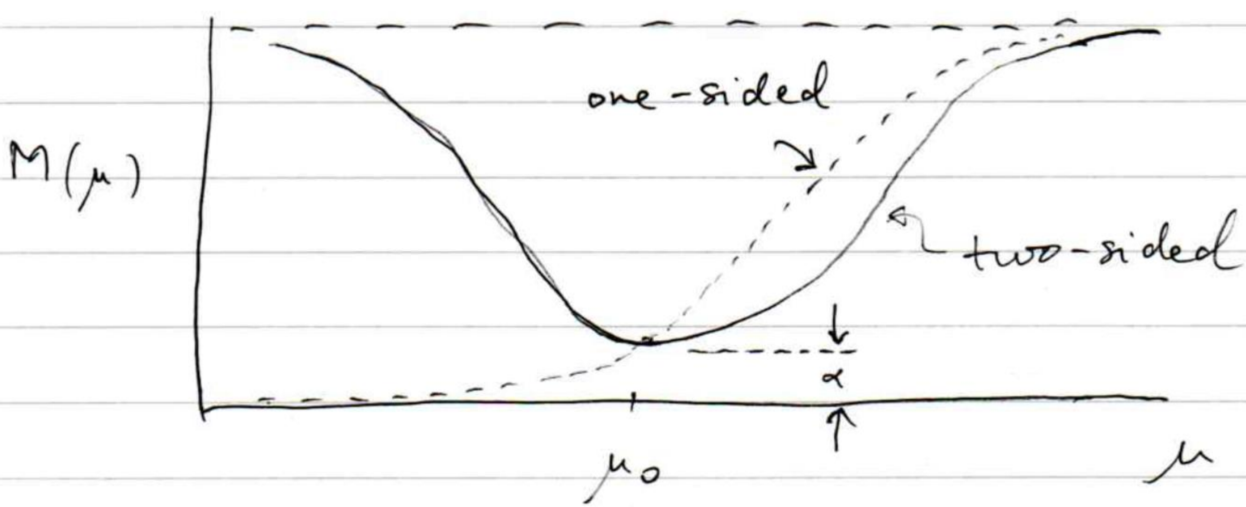


Two-sided w gives (some) power for

both $\mu < \mu_0$ and $\mu > \mu_0$,

but less power for $\mu > \mu_0$ than the original

one-sided test.



# Chain rule for pdfs

Recall conditional pdf

$$f(x_2, x_1) = f(x_2 | x_1)\, f(x_1)$$

Generalize to

$$f(x_n, \ldots, x_1) = f(x_n | x_{n-1}, \ldots, x_1)\, \underbrace{f(x_{n-1}, \ldots, x_1)}$$

$$f(x_{n-1}, \ldots, x_1) = f(x_{n-1} | x_{n-2}, \ldots, x_1)\, \underbrace{f(x_{n-2}, \ldots, x_1)}$$

$$f(x_{n-2}, \ldots, x_1) = f(x_{n-2} | x_{n-3}, \ldots, x_1)\, f(x_{n-3}, \ldots, x_1)$$

$$\vdots$$

Assemble the factors:

$$f(x_n, \ldots, x_1) = \prod_{i=1}^{n} f(x_i \mid \underbrace{x_{i-1}, \ldots, x_1}_{\text{if } i>1, \text{ else, no condition}})$$

e.g. $f(x_3, x_2, x_1) = f(x_3 \mid x_2, x_1)$

$\times f(x_2 \mid x_1)$

$\times f(x_1)$