Notes for lecture #1

Bayes' theorem example

- Suppose you test positive for a rare disease (1 in 10,000 people have the disease on average)
  The probability that the test comes out positive if you have the disease is $0.95 = p(+|H)$.
  The probability that the test comes out positive if you don't have the disease is $0.01 = p(+|\bar{H})$
  What is the probability that you have the disease?

Want to determine

$p(H \mid +)$ = prob. that you have the disease given that you tested +

$$= \frac{p(+ \mid H)\, p(H)}{p(+)}$$

where $p(+ \mid H) = 0.95$

$p(H) = 0.0001$

$p(+) = p(+ \mid H)\, p(H) + p(+ \mid \bar{H})\, p(\bar{H})$

$= 0.95 \times 0.0001 + 0.01 \times 0.999$

$\approx 0.0001 + 0.01$

≈ 0.01

Thus, $p(H \mid +) \approx \dfrac{0.95 \times 0.0001}{0.01}$

$= 0.95 \times 0.01$

≈ 0.01 — so $\frac{1}{100}$ instead of $\frac{1}{10{,}000}$

Example:

$$p(d \mid M_0) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left[-\frac{1}{2\sigma^2}\sum_i d_i^2\right]$$

$$p(d \mid a, M_1) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left[-\frac{1}{2\sigma^2}\sum_i (d_i - a)^2\right]$$

Maximum likelihood estimator of a:

$$0 = \left.\frac{dp}{da}\right|_{a=\hat{a}} = \left. \exp[\;\;]\; -\frac{1}{2\sigma^2}\, 2 \sum_i (d_i - a)\right|_{a=\hat{a}}$$

$$\Longleftrightarrow \quad 0 = \left.\sum_i (d_i - a)\right|_{a=\hat{a}}$$

$$= \sum_i d_i - \hat{a} N$$

$$\rightarrow \quad \boxed{\hat{a} = \frac{1}{N}\sum_i d_i}$$

$$\sum_i (d_i - a)^2 = \sum_i (d_i^2 + a^2 - 2a d_i)$$

$$= \sum_i d_i^2 + N a^2 - 2a \sum_i d_i$$

$$= N\left(\frac{1}{N}\sum_i d_i^2\right) + N a^2 - 2aN\left(\frac{1}{N}\sum_i d_i\right)$$

$$= N\left[\frac{1}{N}\sum_i d_i^2 + a^2 - 2a\hat{a}\right]$$

Now $\mathrm{Var}[d] = \frac{1}{N}\sum_i (d_i - \bar{d})^2$

$$= \frac{1}{N}\sum_i (d_i - \hat{a})^2$$

$$= \frac{1}{N}\sum_i (d_i^2 + \hat{a}^2 - 2\hat{a} d_i)$$

$$= \frac{1}{N}\sum_i d_i^2 + \hat{a}^2 - 2\hat{a}^2$$

$$= \frac{1}{N}\sum_i d_i^2 - \hat{a}^2$$

$$\rightarrow \sum_i (d_i - a)^2 = N\left[\mathrm{Var}[d] + (a - \hat{a})^2\right]$$

thus,

$$p(d|a, M_1) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left[-\frac{1}{2\sigma^2}\sum_i (d_i - a)^2\right]$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left[-\frac{N}{2\sigma^2}\left(\mathrm{Var}[d] + (a-\hat{a})^2\right)\right]$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left[-\frac{\mathrm{Var}[d]}{2\sigma_{\hat{a}}^2}\right] \underbrace{\exp\left[-\frac{1}{2\sigma_{\hat{a}}^2}(a-\hat{a})^2\right]}_{\text{parameter } a \text{ only enters here}}$$

Evidence:

$$p(d|M_1) = \int_0^\infty da\; p(d|a, M_1)\, p(a|M_1)$$

$$p(a|M_1) = \begin{cases} \frac{1}{a_{max}} & 0 \le a \le a_{max} \\ 0 & a_{max} < a \end{cases}$$

$$= \frac{1}{a_{max}} \int_0^{a_{max}} da\; p(d|a, M_1)$$

$$= \frac{1}{a_{max}} \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left[\frac{-\mathrm{Var}[d]}{2\sigma_{\hat{a}}^2}\right] \int_0^{a_{max}} da\; \exp\left[-\frac{1}{2\sigma_{\hat{a}}^2}(a-\hat{a})^2\right]$$

Now:

$$\int_0^{a_{max}} da \, \exp\left[-\frac{1}{2\sigma_{\hat{a}}^2}(a-\hat{a})^2\right] = \int_{-\hat{a}}^{a_{max}-\hat{a}} dx \, \exp\left[-\frac{x^2}{2\sigma_{\hat{a}}^2}\right]$$

Let $x = a - \hat{a}$, $dx = d$

$a = 0, a_{max} \rightarrow x = -\hat{a}, a_{max} - \hat{a}$

Let $t = \frac{x}{\sqrt{2}\sigma_{\hat{a}}} \rightarrow dt = \frac{dx}{\sqrt{2}\sigma_{\hat{a}}}$

$$= \int_{\frac{-\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}}^{(a_{max}-\hat{a})/\sqrt{2}\sigma_{\hat{a}}} dt \, \sqrt{2}\sigma_{\hat{a}} \, e^{-t^2}$$

$$= \sqrt{2}\sigma_{\hat{a}} \int_{\frac{-\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}}^{\frac{a_{max}-\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}} dt \, e^{-t^2}$$

$$= \sqrt{2}\sigma_{\hat{a}} \left[\int_{\frac{-\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}}^{0} + \int_0^{(a_{max}-\hat{a})/\sqrt{2}\sigma_{\hat{a}}}\right] dt \, e^{-t^2}$$

Write in terms of $\mathrm{erf}(z)$:

$$\mathrm{erf}(z) \equiv \frac{2}{\sqrt{\pi}}\int_0^z dt \, e^{-t^2}$$

$$\int_0^{a_{max}} da \, \exp\left[-\frac{1}{2\sigma_{\hat{a}}^2}(a-\hat{a})^2\right] = \sqrt{2}\,\sigma_{\hat{a}} \left[\int_0^{(a_{max}-\hat{a})/\sqrt{2}\sigma_{\hat{a}}} - \int_0^{-a_{max}/\sqrt{2}\sigma_{\hat{a}}}\right] dt\, e^{-t^2}$$

$$= \sqrt{2}\,\sigma_{\hat{a}} \frac{\sqrt{\pi}}{2}\left[\operatorname{erf}\left(\frac{a_{max}-\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right) + \operatorname{erf}\left(\frac{\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right)\right]$$

thus,

$$p(d|M_1) = \frac{1}{a_{max}}\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left(-\frac{Var[d]}{2\sigma_{\hat{a}}^2}\right)\frac{\sigma_{\hat{a}}\sqrt{\pi}}{\sqrt{2}}\left[\operatorname{erf}\left(\frac{a_{max}-\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right) + \operatorname{erf}\left(\frac{\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right)\right]$$

$$= \frac{\exp\left(-\frac{Var[d]}{2\sigma_{\hat{a}}^2}\right)\left[\operatorname{erf}\left(\frac{a_{max}-\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right) + \operatorname{erf}\left(\frac{\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right)\right]}{\sqrt{2}\,(\sqrt{2\pi})^N \sigma^N \frac{\sqrt{N}}{\sigma\sqrt{\pi}}\, a_{max}}$$

$$= \boxed{\frac{\exp\left(-\frac{Var[d]}{2\sigma_{\hat{a}}^2}\right)\left[\operatorname{erf}\left(\frac{a_{max}-\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right) + \operatorname{erf}\left(\frac{\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right)\right]}{2\,a_{max}\,(\sqrt{2\pi}\,\sigma)^{N-1}\sqrt{N}}}$$

Posterior distribution

$$p(a \mid d, M_1) = \frac{p(d \mid a, M_1)\, p(a \mid M_1)}{p(d \mid M_1)}$$

$$= \frac{\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \cancel{\exp\left(-\frac{Var[d]}{2\sigma_{\hat{a}}^2}\right)} \exp\left(-\frac{1}{2}\frac{(a-\hat{a})^2}{\sigma_{\hat{a}}^2}\right) \cancel{\frac{1}{a_{max}}}}{\frac{\cancel{\exp\left(\frac{-Var[d]}{2\sigma_{\hat{a}}^2}\right)} \left[\mathrm{erf}(\ ) + \mathrm{erf}(\ )\right]}{2 \cancel{a_{max}} (\sqrt{2\pi}\,\sigma)^{N-1} \sqrt{N}}}$$

$$= \frac{1}{\sqrt{2\pi}}\frac{\sqrt{N}}{\sigma} \exp\left(-\frac{1}{2}\frac{(a-\hat{a})^2}{\sigma_{\hat{a}}^2}\right) 2\left[\mathrm{erf}(\,) + \mathrm{erf}(\,)\right]^{-1}$$

$$\boxed{p(a \mid d, M_1) = \frac{1}{\sqrt{2\pi}\,\sigma_{\hat{a}}} \exp\left(-\frac{1}{2}\frac{(a-\hat{a})^2}{\sigma_{\hat{a}}^2}\right) 2\left[\mathrm{erf}\left(\frac{a_{max}-\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right) + \mathrm{erf}\left(\frac{\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right)\right]^{-1}}$$

Truncated gaussian on $[0, a_{max}]$

Bayes' Factor:

$$B_{10}(d) \equiv \frac{p(d|M_1)}{p(d|M_0)}$$

doesn't have any free parameters

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left[-\frac{\sum_i d_i^2}{2\sigma^2}\right]$$

$$= \frac{\exp\left(-\frac{Var[d]}{2\sigma_{\hat{a}}^2}\right) \frac{1}{2a_{max}\,(\sqrt{2\pi}\,\sigma)^{N-1}\sqrt{N}} \left[erf(\;) + erf(\;)\right]}{\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left(-\frac{\sum_i d_i^2}{2\sigma^2}\right)}$$

cancels

Recall: $Var[d] = \frac{1}{N}\sum_i d_i^2 - \hat{a}^2$

$$\rightarrow \exp\left(-\frac{Var[d]}{2\sigma_{\hat{a}}^2}\right) = \exp\left(-\frac{1}{2N\sigma_{\hat{a}}^2}\sum_i d_i^2\right)\exp\left(+\frac{\hat{a}^2}{2\sigma_{\hat{a}}^2}\right)$$

$$= \exp\left(-\frac{\sum_i d_i^2}{2\sigma^2}\right)\exp\left(\frac{\hat{a}^2}{2\sigma_{\hat{a}}^2}\right)$$

$$B_{10}(d) = \frac{\exp\left(\frac{\hat{a}^2}{2\sigma_{\hat{a}}^2}\right)\left[\mathrm{erf}(\,) + \mathrm{erf}(\,)\right]}{2a_{max}} \frac{\left(\sqrt{2\pi}\,\sigma\right)^N}{\left(\sqrt{2\pi}\,\sigma\right)^{N-1}\sqrt{N}}$$

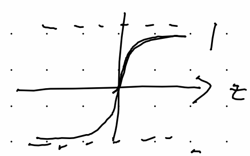
$$= \exp\left(\frac{\hat{a}^2}{2\sigma_{\hat{a}}^2}\right)\frac{\left[\mathrm{erf}(\,) + \mathrm{erf}(\,)\right]}{2}\left(\frac{\sqrt{2\pi}\;\sigma_{\hat{a}}}{a_{max}}\right)$$

$$= \exp\left(\frac{\hat{a}}{2\sigma_{\hat{a}}^2}\right)\left(\frac{\sqrt{2\pi}\;\sigma_{\hat{a}}}{a_{max}}\right)\frac{1}{2}\left[\mathrm{erf}\left(\frac{a_{max}-\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right) + \mathrm{erf}\left(\frac{\hat{a}}{\sqrt{2}\,\sigma_{\hat{a}}}\right)\right]$$

Note:

$$B_{10}(d) \simeq \exp\left(\frac{\hat{a}^2}{2\sigma_{\hat{a}}^2}\right)\left(\frac{\sqrt{2\pi}\,\sigma_{\hat{a}}}{a_{max}}\right)$$

If $\hat{a}$ tightly peaked away from 0 and a_{max}



erf(z)

Note: $\Lambda(d)$ = frequentist detection statistic

$$= 2 \ln \Lambda_{ML}(d)$$

$$= 2 \ln \left[\frac{p(d \mid a, \mathcal{M}_1) \big|_{a = a_{ML}}}{\underbrace{p(d \mid \mathcal{M}_0)}_{\text{no parameters}}} \right]$$

$$p(d \mid a, \mathcal{M}_1) = \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^N \exp\left(- \frac{\mathrm{Var}[d]}{2 \sigma_{\hat a}^2} \right) \exp\left(- \frac{(a - \hat a)^2}{2 \sigma_{\hat a}^2} \right)$$

$$\longrightarrow \quad p(d \mid a = \underbrace{\hat a}_{\text{max likelihood estimator}}) = \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^N \exp\left(\frac{-\mathrm{Var}[d]}{2 \sigma_{\hat a}^2} \right)$$

Thus,

$$\Lambda_{ML}(d) = \frac{\cancel{\left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^N} \exp\left(- \frac{\mathrm{Var}[d]}{2 \sigma_{\hat a}^2} \right)}{\cancel{\left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^N} \exp\left(- \frac{\sum_i d_i^2}{2 \sigma^2} \right)}$$

$$\Lambda_{ML}(d) = \frac{\exp\left(-\frac{\mathrm{Var}[d]}{2\sigma_{\hat{a}}^2}\right)}{\exp\left(-\frac{\sum_i d_i^2}{2\sigma^2}\right)}$$

Recall: $\mathrm{Var}[d] = \frac{1}{N}\sum_i d_i^2 - \hat{a}^2$

$$\rightarrow \quad \exp\left(-\frac{\mathrm{Var}[d]}{2\sigma_{\hat{a}}^2}\right) = \exp\left(\frac{-\sum_i d_i^2}{2N\sigma_{\hat{a}}^2}\right)\exp\left(\frac{+\hat{a}^2}{2\sigma_{\hat{a}}^2}\right)$$

$$= \exp\left(\frac{-\sum_i d_i^2}{2\sigma^2}\right)\exp\left(\frac{\hat{a}^2}{2\sigma_{\hat{a}}^2}\right)$$

so

$$\Lambda_{ML}(d) = \exp\left(\frac{\hat{a}^2}{2\sigma_{\hat{a}}^2}\right)$$

$$\Lambda(d) \equiv 2\ln\Lambda_{ML}(d) = \frac{\hat{a}^2}{\sigma_{\hat{a}}^2} = \boxed{\frac{N\hat{a}^2}{\sigma^2}} \leftarrow \text{squared SNR}$$

For informative data:

$$B_{10}(d) \simeq \exp\left(\frac{+\hat{a}^2}{2\sigma_{\hat{a}}^2}\right)\left(\frac{\sqrt{2\pi}\,\sigma_{\hat{a}}}{a_{max}}\right)$$

$$2\ln B_{10}(d) \simeq \frac{\hat{a}}{\sigma_{\hat{a}}^2} + 2\ln\left(\frac{\sqrt{2\pi}\,\sigma_{\hat{a}}}{a_{max}}\right)$$

$$= 2\ln\Lambda_{ML}(d) + 2\ln\left(\frac{\sqrt{2\pi}\,\sigma_{\hat{a}}}{a_{max}}\right)$$

$$= \Lambda(d) + 2\ln\underbrace{\left(\frac{\sqrt{2\pi}\,\sigma_{\hat{a}}}{a_{max}}\right)}_{\text{Occam Factor} \sim \frac{\Delta V_1}{V_1}}$$

Sampling distribution of frequentist detection statistic

$$\Lambda(d) = \frac{N\hat{a}^2}{\sigma^2} = \left(\frac{\sqrt{N}\,\bar{d}}{\sigma}\right)^2 \equiv \rho^2 \quad \text{where } \rho \equiv \frac{\sqrt{N}\,\bar{d}}{\sigma}$$

Now: ρ is gaussian distributed being an average of d_i ← gauss

(i) In the absence of a signal: $\langle \rho \rangle = 0$

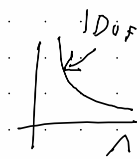
$$\mathrm{Var}(\rho) = \frac{N}{\sigma^2}\underbrace{\mathrm{Var}(\bar{d})}_{\sigma^2/N} = 1$$

(ii) In the presence of a signal: $\langle \rho \rangle = \frac{\sqrt{N}\,a}{\sigma} \equiv \mu$

$\mathrm{Var}(\rho) = 1$ (since signal is deterministic)

Central chi-square with 1 DOF:

$$p(\Lambda \,|\, \mathcal{M}_0) = \frac{1}{\sqrt{2}\,\underbrace{\Gamma(\tfrac{1}{2})}_{\pi}}\,\Lambda^{-1/2}\,e^{-\Lambda/2} = \boxed{\frac{1}{\sqrt{2\pi\Lambda}}\,e^{-\Lambda/2}}$$



$p(\Lambda \mid a, M_1)$ = non-central chi-square distribution with one DOF $\lambda = \mu^2 = \frac{Na^2}{\sigma^2}$

$$= \frac{1}{2} e^{-(\Lambda+\lambda)/2} \left(\frac{\Lambda}{\lambda}\right)^{-\frac{1}{4}} \cdot I_{-\frac{1}{2}}\left(\sqrt{\lambda\Lambda}\right)$$

$$= \frac{1}{2\sqrt{\Lambda}} \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{\Lambda}-\sqrt{\lambda})^2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{\Lambda}+\sqrt{\lambda})^2} \right]$$

$$= \boxed{\frac{1}{\sqrt{2\pi\Lambda}} \frac{1}{2} \left[e^{-\frac{1}{2}(\sqrt{\Lambda}-\sqrt{\lambda})^2} + e^{-\frac{1}{2}(\sqrt{\Lambda}+\sqrt{\lambda})^2} \right]}$$

for $0 \le \Lambda < \infty$

where $\lambda = \mu^2 = \frac{Na^2}{\sigma^2}$