# Lecture 3

## Hypothesis testing:

- hypothesis $H$: model claiming to explain the observation

  $H$ can be i) simple : fully defined

  ii) composite: has free parameter(s)

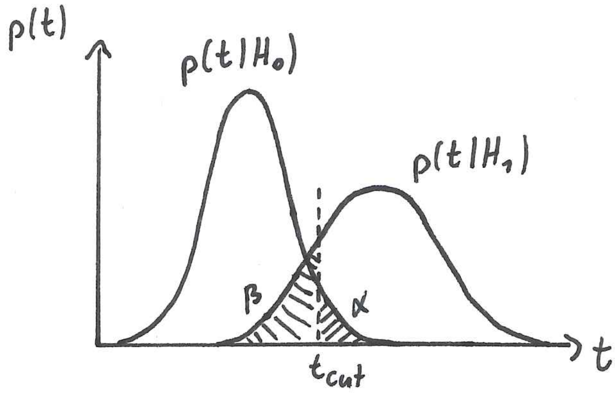
- hypothesis testing: deciding between two different $H_i$

  $H_0$ null hypothesis (SM-only in case of discovery)

  $H_1$ alternative hypothesis (SM + signal in case of discovery)

- test statistic $t$ : · (usually scalar) variable determined from data & $H$
  - · should distinguish between $H_0$ & $H_1$
  - · e.g. $\chi^2$, neural network score, ...

- define $t_{cut}$ such that you choose $\begin{cases} H_1 & \text{if } t_{obs} > t_{cut} \\ H_0 & \text{else} \end{cases}$




- probabilities for wrong choice:

$$\alpha \equiv \int_{t_{cut}}^{\infty} p(t|H_0)\, dt$$ false-positive rate, type-I error, "size of test"

$1-\alpha$ : specificity

$$\beta \equiv \int_{-\infty}^{t_{cut}} p(t|H_1)\, dt$$ false-negative rate, type-II error

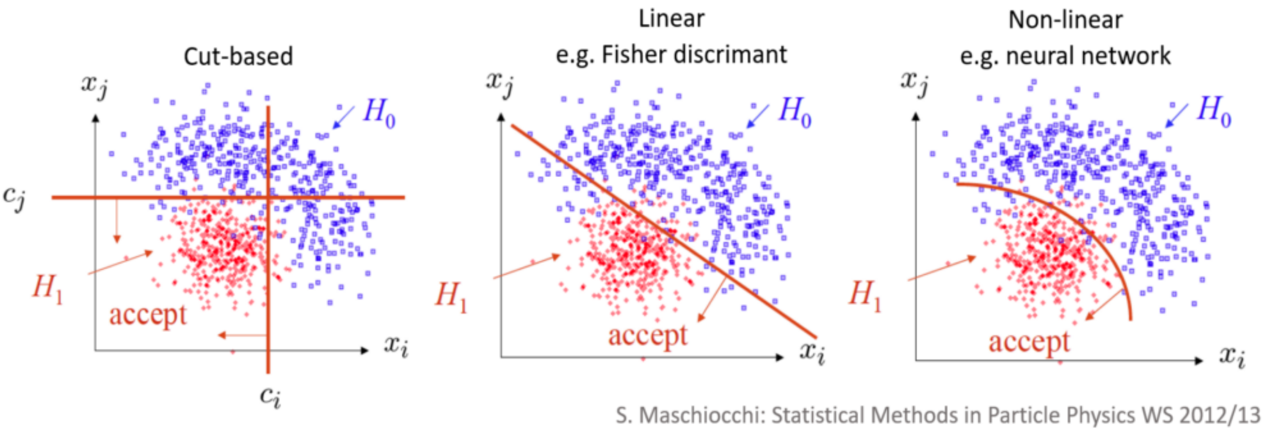
$1-\beta$ : "power of test", sensitivity

• multivariate: • define critical region $W$ instead

$$p(\vec{x} \in W \mid H_0) \leq \alpha$$

$$p(\vec{x} \in W \mid H_1) = \beta$$

• e.g. cut-based, BDT, $\underbrace{\text{Fisher discriminant, neural network}}_{\text{multivar} \to \text{scalar}}$, ...




S. Maschiocchi: Statistical Methods in Particle Physics WS 2012/13

• choice of test statistic:

Neyman-Pearson lemma:

• given simple $H_0, H_1$, significance level $\alpha$, the most powerful test (smallest $\beta$)

is the likelihood ratio: (or a monotonic fct. of it)

$$t(\vec{x}) = \frac{\mathcal{L}(H_1 \mid \vec{x})}{\mathcal{L}(H_0 \mid \vec{x})}$$

conditional MLE: best fit NPs for fix $\mu$

• at LHC we use: $\lambda_\mu \equiv \dfrac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}_\mu \mid \vec{x})}{\underbrace{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}} \mid \vec{x})}_{=\mathcal{L}_{max}}}$

test stat. for discovery: $q_0 \equiv \begin{cases} -2 \ln \lambda_0 & \text{if } \hat{\mu} \geq 0 \\ 0 & \text{else} \end{cases}$

test stat. for limits: $q_\mu \equiv \begin{cases} -2 \ln \lambda_\mu & \text{if } \mu \geq \hat{\mu} \geq 0 \\ 0 & \text{else} \end{cases}$

• to determine $q_\mu^{cut}$ corresponding to CL $1-\alpha$, we must know $p(q_\mu)$!

i) MC method

• fix NPs to best fit $\hat{\hat{\vec{\theta}}}$, generate pseudodata

• histogram resulting $q_\mu$'s $\rightarrow p(q_\mu)$

ii) asymptotic approximation

- Wald's theorem: $-2\ln\lambda_\mu \simeq \frac{(\mu-\hat{\mu})^2}{\sigma^2}$ in large sample limit
- Cowan et al. (2010): analytic formulas for $p(q_0)$, $p(q_\mu)$:

$$p(q_0|\mu') = \left(1-\Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0}-\frac{\mu'}{\sigma}\right)^2\right]$$

$$p(q_\mu|\mu') = \Phi\left(\frac{\mu'-\mu}{\sigma}\right)\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{q_\mu}-\frac{\mu-\mu'}{\sigma}\right)^2\right]$$

- they depend on std. dev. $\sigma$ of estimator $\hat{\mu}$ (assumed Gaussian around true $\mu'$) estimated via "Asimov data set":
  - hypothetical data set where estimators equal the true values $\hat{\mu}=\mu'$, $\hat{\vec{\theta}}=\vec{\theta}'$
  - approximated by nominal SM-only MC ($\mu'=\hat{\mu}=0$)

  $\Rightarrow q_\mu^A \simeq \frac{\mu^2}{\sigma^2}$ determines $\sigma$

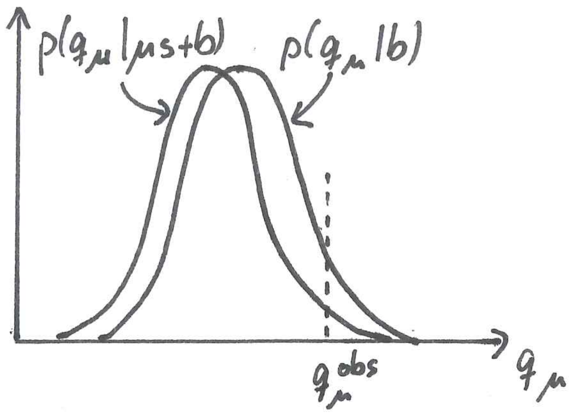
$\rightarrow$ with $p(q_\mu)$, $p(q_0)$ we can now determine p-values:

<u>Discovery:</u> • $p_0 = P(q_0 \geq q_0^{obs}|b) = \int_{q_0^{obs}}^{\infty} p(q_0|\mu=0,\hat{\hat{\theta}}_0)\,dq_0$

- claim discovery if $p_0 < 2.9\times10^{-7}$ ($\hat{=}\,5\sigma$)
- this is only local p-value, if testing many hypotheses account for look-elsewhere effect! (e.g. by including trials-factor)

<u>Exclusion limits:</u> • $CL_{s+b}(\mu) \equiv p_\mu = P(q_\mu \geq q_\mu^{obs}|\mu s+b) = \int_{q_\mu^{obs}}^{\infty} p(q_\mu|\mu,\hat{\hat{\theta}}_\mu)\,dq_\mu$

- exclude signal strength $\mu$ if $CL_{s+b}(\mu) < 0.05$
- problem: $CL_{s+b}$ will exclude any model in 5% of analyses
  $\rightarrow$ most likely underfluctuations of bkg




- solution: "punish" insensitive experiments $\rightarrow CL_s$

"Despite its shaky foundations in statistical theory, it has been producing sensible results for over a decade." - A. Harel (CMS)

- also consider p-value of bkg-hypothesis:

$$CL_b(\mu) \equiv 1-p_b = P(q_\mu \geq q_\mu^{obs}|b) = \int_{q_\mu^{obs}}^{\infty} p(q_\mu|\mu=0,\hat{\hat{\theta}}_0)\,dq_\mu$$

$$CL_s(\mu) \equiv \frac{CL_{s+b}(\mu)}{CL_b(\mu)}$$

- $CL_s$ is no longer a p-value, but we treat it as one:

  exclude signal strength $\mu$ if $CL_s(\mu) < 0.05$

- $CL_s$ is always more conservative than $CL_{s+b}$ because $0 < CL_b \leq 1$

- upper limit $\mu^{95\%\,CL}$ by inverting hypothesis test:

  find intersection of $CL_s(\mu)$ with $CL_s = 0.05$