Overview of Bayesian calculation:

A) Define the problem

Goal: Get the joint distribution of everything random & relevant.
Method: Define notation, factor joint distribution, calculate each part.

Notation:

Role	What is known	Unknown, we *don't* care about it	Unknown, we *do* care about it
Symbol	X	ψ	θ
Interpretation	Observed data; anything else that is known.	Nuisance parameter, Choice of model, Choice of prior	Parameter, future observable, asymptotic observable

Joint distribution: $[X,\psi,\theta] \quad = \quad [\psi,\theta][X\,|\,\psi,\theta]$

The first term $[\psi,\theta]$ describes prior knowledge.

The second term $[X\,|\,\psi,\theta]$ describes the mechanism generating the data.

B) Apply Bayes "theorem": Calculate the (posterior) distribution of the quantity of interest:

Role	What is known	Unknown, we *don't* care about it	Unknown, we *do* care about it
Symbol	X	ψ	θ
What to do with it	Condition on it!	Integrate it out!	Learn from the posterior distrib'n! Make a decision!

$$posterior = \frac{model \cdot prior}{marginal}$$

$$[\theta\,|\,X] = \frac{[X\,|\,\theta]\cdot[\theta]}{[X]}$$ (no nuisance ψ)

$$[\theta\,|\,X] = \frac{\int_{\psi} [X,\psi,\theta]}{\int_{\psi,\theta} [X,\psi,\theta]}$$ (nuisance ψ is present)

IMPORTANT: the posterior ONLY depends on the likelihood function. The sampling mechanism (which *h*() reflects) does not affect the result.

$$[\theta\,|\,X] \quad = \quad \frac{h(X)[X\,|\,\theta][\theta]}{\int_{\theta\in\Theta} h(X)[X\,|\,\theta][\theta]} = \frac{[X\,|\,\theta][\theta]}{\int_{\theta\in\Theta} [X\,|\,\theta][\theta]}$$

regardless of *h*(). This is not true in classical statistics.

C) *Decision-making:*

Goal: Given a posterior distribution [θ | X], choose the best action to take.

Method: comparing Bayes expected loss for various actions:

C.1. Specify the ***loss function*** $L : \mathcal{A} \times \Theta \rightarrow \Re$, so that $L(\theta, a)$ for θ in Θ and a in A (actions a, action space A) represents the loss ensuing if you take action a when the true state of nature is θ.

C.2 Calculate the ***Bayes expected loss***:

$$\rho(a \mid X) = E_{\theta|X}(L(\theta, a)) = \int_{\theta} L(\theta, a)[\theta \mid X]$$

C.3 Choose ***Bayes action***,

$$a_B = \arg\min \rho(a \mid X)$$

All this is done after taking account all current knowledge (both current data and prior).

An entire PLAN for making decisions is called a ***decision rule***:

$$\delta : \mathcal{X} \rightarrow \mathcal{A}$$

For each data set $X \in \mathcal{X}$, our plan tells us what action $a \in \mathcal{A}$ to take.

A decision rule that agrees with Bayes action for each X is called a ***Bayes rule***.

$$\delta_B : \mathcal{X} \rightarrow \mathcal{A}$$

Each $\delta_B(X)$ is a Bayes action.

Three kinds of expected loss

Averaging over the posterior, $[\theta \mid X]$:

Bayes expected loss: $\rho(a \mid X) = E_{\theta|X}(L(\theta, a))$

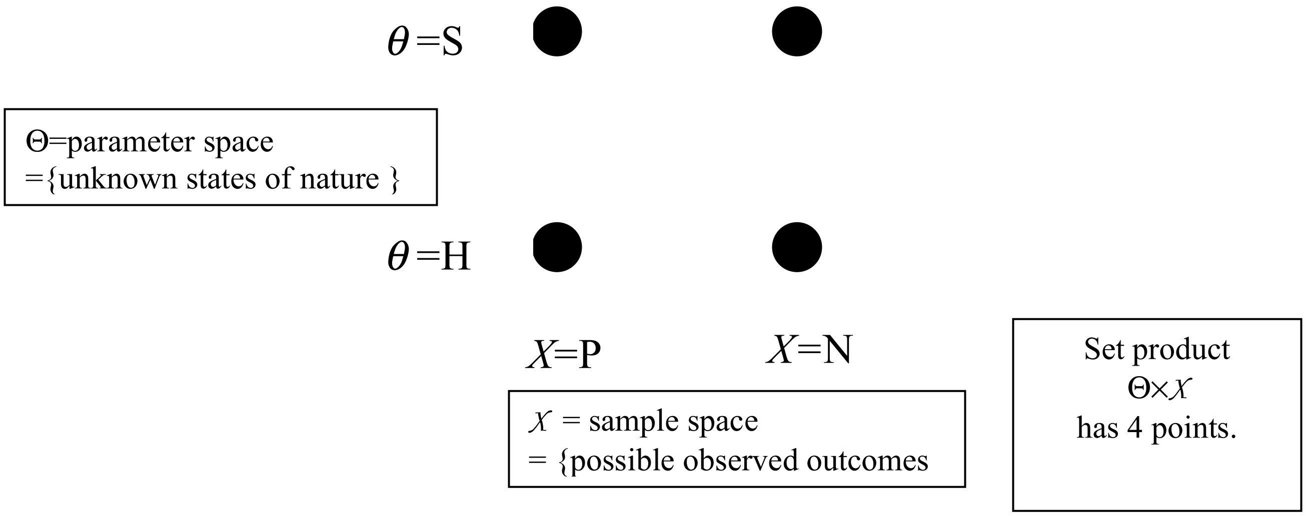
Averaging over the model, $[X \mid \theta]$:

Risk function: $R(\theta;\ \delta, L) = E_{X|\theta}(L(\theta, \delta(X)))$.

Averaging over the joint distribution, $[\theta, X]$:

Bayes risk: $r(\delta; \pi, L) = E_{X,\theta}(L(\theta, \delta(X)))$

Example: unknown is in $\Theta = \{S,H\}$, data is in $\mathcal{X} = \{P,N\}$, action is in $\mathcal{A} = \{\text{Treat},\text{Wait}\}$.



Interpretation:

θ=H "patient is healthy" θ=S "patient is sick"

X=P "test is positive" X=N "test is negative"

A=T "decide to treat" A=W "decide to wait"

Prior = disease prevalence = Pr (θ=S)

Model = $\begin{cases} \text{Sensitivity} = \Pr(X{=}P \mid \theta{=}S) \\ \text{Specificity} = \Pr(X{=}N \mid \theta{=}H) \end{cases}$

Posterior = $\begin{cases} \text{Predictive value of P} = \Pr(\theta{=}S \mid X{=}P) \\ \text{Predictive value of N} = \Pr(\theta{=}H \mid X{=}N) \end{cases}$

$[X,\theta]$	=	$[\theta]$	$[X \mid \theta]$	=	$[X]$	$[\theta \mid X]$
joint	=	marg'l	cond'l	=	marg'l	cond'l
joint	=	prior	model	=	normalizer	posterior

Loss table $\mathcal{A} \times \Theta \rightarrow \Re$ where $\mathcal{A} = \{ \mathbf{T}, \mathbf{W} \}$

	T (treat the patient)	**W (wait)**
Healthy ("null hypothesis")	Loss(T,H) = L_{TH}	Loss(W,H) = 0
Sick ("alternate hypothesis")	Loss(T,S) = 0	Loss(W,S) = L_{WS}

Bayes expected loss:

$$E(\text{Loss} \mid P, T) = L_{TH} \Pr(H \mid P) = L_{TH} \Pr(P \mid H) \Pr(H) / \Pr(P)$$
$$E(\text{Loss} \mid P, W) = L_{WS} \Pr(S \mid P) = L_{WS} \Pr(P \mid S) \Pr(S) / \Pr(P)$$

Choose T over W ("the Bayes action is T") if

$$E(\text{Loss} \mid P, W) > E(\text{Loss} \mid P, T), \quad \text{i.e.}$$

$$L_{WS} \Pr(P \mid S) \Pr(S) / \Pr(P) > L_{TH} \Pr(P \mid H) \Pr(H) / \Pr(P)$$

$$\frac{L_{WS}}{L_{TH}} \; \frac{\Pr(P \mid S)}{\Pr(P \mid H)} \; \frac{\Pr(S)}{\Pr(H)} > 1$$

{loss ratio} {likelihood ratio} {prior odds } > 1

Example:

1) If prevalence = 0.001, sensitivity = 0.99, specificity = 0.99, loss ratio=1,
   and a "P" test result is observed,
2) then prior odds $\approx 1000^{-1}$, likelihood ratio $\approx$ 100,
3) so the Bayes action is ***W*** ("watchful waiting"; do nothing for now).
   [What is the posterior odds?]

   [A "veritasium" video explains this example in detail: https://youtu.be/R13BD8qKeTg , and shows "Bayesian updating" as new data arrive.
   But he neglects what to do if the "conditional independence" assumption is wrong.]

A test splits a group of patients into two, *Neg* and *Pos*; when is it a USEFUL test?

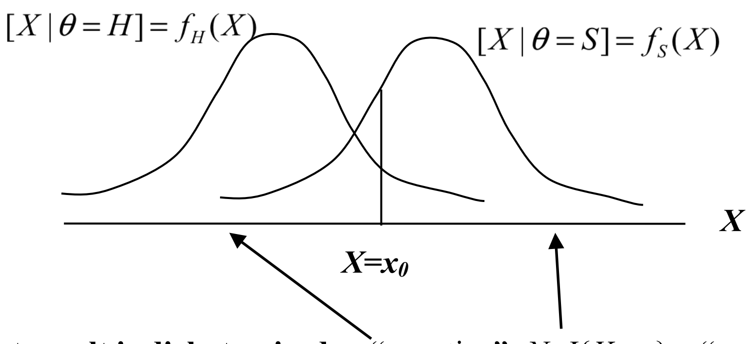
Question: When will the Bayes action be different in the two groups?
Answer:

$$\frac{1}{NPO} = \frac{Pr(S \mid Neg)}{Pr(H \mid Neg)} = \frac{Pr(S)}{Pr(H)} \frac{\Pr(Neg \mid S)}{\Pr(Neg \mid H)} < \frac{L_{Treat,H}}{L_{Wait,S}} <= \frac{Pr(S)}{Pr(H)} \frac{\Pr(Pos \mid S)}{\Pr(Pos \mid H)} = \frac{Pr(S \mid Pos)}{Pr(H \mid Pos)} = PPO$$

where the negative and positive predictive odds are: $NPO = \frac{\Pr(H \mid Neg)}{\Pr(S \mid Neg)}$, $PPO = \frac{\Pr(S \mid Pos)}{\Pr(H \mid Pos)}$.

<u>Setting: diagnostic or screening test with continuous values.</u>

Goal: choose the *threshold*, or *cutoff,* for taking the action.

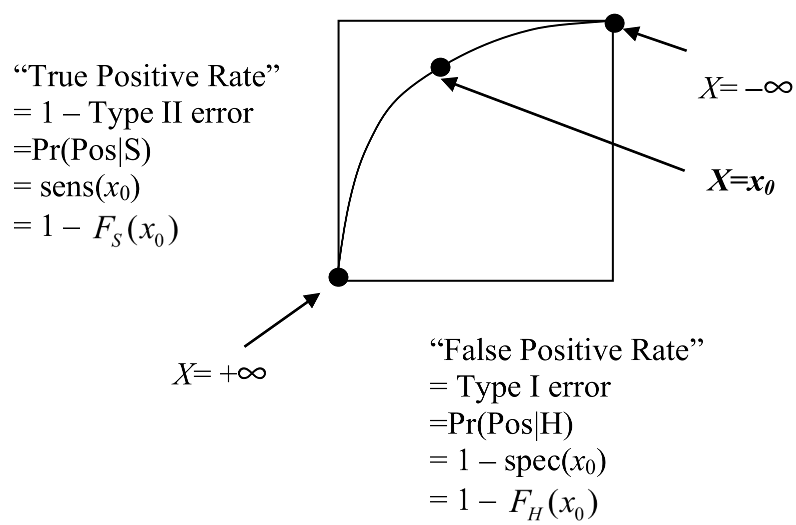


Test result is dichotomized: "negative"=$N=I\{X<x_0\}$, "positive"=$P=I\{X>x_0\}$.

Loss table $\mathcal{A} \times \Theta \to \Re$, $\mathcal{A}$ **= { T, W }**

	T (take an action)	**W (wait)**
Healthy ("null hypothesis")	Loss(T,H) = L_{TH}	Loss(W,H) = 0
Sick ("alternate hypothesis")	Loss(T,S) = 0	Loss(W,S) = L_{WS}

<u>"ROC curve" (receiver operating characteristic)</u>



$$\rho(A \mid X) - \rho(W \mid X) = (L_{AH} \Pr(H \mid X) + 0) - (0 + L_{WS} \Pr(S \mid X))$$
$$= L_{AH} \frac{f_H(X)(1-\pi)}{f_H(X)(1-\pi) + f_S(X)\pi} - L_{WS} \frac{f_S(X)(\pi)}{f_H(X)(1-\pi) + f_S(X)\pi}$$

This is positive if

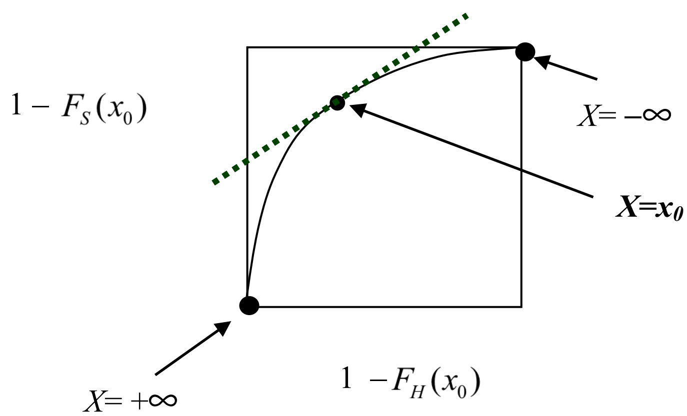
$$\frac{L_{AH}}{L_{WS}} \frac{f_H(X)}{f_S(X)} \frac{(1-\pi)}{\pi} > 1.$$

The optimal cutoff x_0 satisfies

$$\frac{L_{AH}}{L_{IS}} \frac{f_H(x_0)}{f_S(x_0)} \frac{(1-\pi)}{\pi} = 1$$

$$\text{Slope} = \frac{\partial(1-F_S(x_0))/\partial x}{\partial(1-F_H(x_0))/\partial x} = \frac{f_S(x_0)}{f_H(x_0)} = \frac{L_{AH}}{L_{WS}} \frac{(1-\pi)}{\pi}$$

common sense check?



(A cute aside: pick one S person and one H person. The *X* value for the S person should be larger—usually. What is the probability that this is true? Think about ROC graph.)

EXERCISE: if the two distributions are normal, what is the likelihood ratio, between the hypotheses "S" versus "H", if the variances are the same? different?