Base Rates and Bayes' Theorem

Slides to accompany Grove's handout

Scott Vrieze

March 8, 2016

Table of contents

Example inferences

- Does this patient have Alzheimer's disease, schizophrenia, depression, etc.?
- What is the patient's cognitive ability?
- How amenable is this patient to talk therapy?
- Was this individual abused as a child?

Pieces of information

- **Symptoms** Self-report measures obtained via interview, whether structured or not
- **Behavioral observations and other signs** Signs are observable characteristics, in contrast to symptoms
- **Life history facts**
- **Psychological test results such as scores or unscored data**
- **Physiological measures**
- **Others?**

Fundamental Quantities

- X is the observable characteristic (sadness, defense mechanism, etc.)
- Y is the latent state (depression, schizophrenia, etc.)
- Π_s and Π_n are two non-overlapping subpopulations
- P = Population point prevalence of Π_s
- $Q = 1 - P$
- α is the sensitivity, or $\Pr\{X = 1|Y = 1\}$. (What is $1 - \alpha$?)
- β is the specificity, or $\Pr\{X = 0|Y = 0\}$. (What is $1 - \beta$?)

Bayes' Theorem

- α and β give information about one's score on the test *given their diagnostic status*
- We want to go the other way, and infer diagnostic status *given their test score*
- Bayes' theorem let's us do just that.

$$\Pr\{Y|X\} = \frac{\Pr\{X|Y\}\Pr\{Y\}}{\Pr\{X\}},$$

thus the probabilty of having the disorder given a positive test score is:

$$\Pr\{Y=1|X=1\} = \frac{\Pr\{X=1|Y=1\}\Pr\{Y=1\}}{\Pr\{X=1\}},$$

where

- $\Pr\{Y=1\} = P$ is the base rate of the disorder,
- $\Pr\{X=1\}$ is the probability of a positive test score, and
- $\Pr\{X=1|Y=1\} = \alpha$ is the sensitivity.

Derived Quantities: PPV & NPV

Positive predictive value: Probability one is called a case, given one scores positive on the test.

$$\begin{aligned} PPV &= \Pr\{Y = 1|X = 1\} \\ &= \frac{\Pr\{X = 1|Y = 1\}\Pr\{Y = 1\}}{\Pr\{X = 1\}} \\ &= \frac{P\alpha}{P\alpha + Q(1-\beta)} \end{aligned} \tag{1}$$

Negative predictive value: Probability one is called a noncase given one scores negative on the test.

$$\begin{aligned} NPV &= \Pr\{Y = 0|X = 0\} \\ &= \frac{\Pr\{X = 0|Y = 0\}\Pr\{Y = 0\}}{\Pr\{X = 0\}} \\ &= \frac{Q\beta}{P(1-\alpha) + Q\beta} \end{aligned} \tag{2}$$

Derived Quantities: Hit Rate (aka Efficiency)

Hit Rate: Proportion of individuals correctly classified.

$$HR = P\alpha + Q\beta$$

Likelihood and Likelihood Ratio

Probability mass function:

$$g(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Probability density function:

$$g(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- These are probability functions when we consider the parameters fixed and the observed scores x random.
- They are likelihood functions when the observed scores x are fixed (e.g., we collected data) and the parameters are random.

Likelihood and Likelihood Ratio

- Generally, a likelihood function's value for the *i*th patient from population Π_s with test result $X = x_i$ equals the probability function for Π_s at $X = x_i$.
- More specific to symptoms, the likelihood that a randomly chosen individual has symptom X given they are a case is α, by definition.
- The likelihood they have symptom X if they are a noncase is $(1 - \beta)$.
- The quotient $\frac{\alpha}{(1-\beta)}$ is the likelihood ratio, the strength of evidence in favor of the hypothesis that a patient who has symptom X has disorder Y. More formally:

$$\Omega(x_i) = \frac{\Pr\{X = x_i | Y = 1\}}{\Pr\{X = x_i | Y = 0\}}$$

Likelihood and Likelihood Ratio

- The generic LR is symbolized as:

$$\Omega(X = x_i) = \frac{\Pr\{X = x_i | Y = 1\}}{\Pr\{X = x_i | Y = 0\}}$$

- For patients with $X = 1$ the LR is:

$$\Omega_1 = \frac{\alpha}{(1 - \beta)}$$

- For patients with $X = 0$ the LR is:

$$\Omega_0 = \frac{1 - \alpha}{\beta}$$

Likelihood and Likelihood Ratio

- The generic LR is symbolized as:

$$\Omega(X = x_i) = \frac{\Pr\{X = x_i | Y = 1\}}{\Pr\{X = x_i | Y = 0\}}$$

- For patients with $X = 1$ the LR is:

$$\Omega_1 = \frac{\alpha}{(1 - \beta)}$$

- For patients with $X = 0$ the LR is:

$$\Omega_0 = \frac{1 - \alpha}{\beta}$$

Stepwise revision of probabilistic knowledge.

$$
\begin{aligned}
O_{post} &= O_{prior}\Omega_{X=x_i} \\
&= \frac{P}{Q} \times \begin{cases} \frac{\alpha}{(1-\beta)}, & \text{if } X = 1 \\ \frac{(1-\alpha)}{\beta}, & \text{if } X = 0 \end{cases}
\end{aligned}
$$

Worked Example – Clinic A.

In clinic A $P = .01$. Assume we have a patient with a positive test for the disease. Let the sensitivity and specificity be:

- $\alpha = .875$
- $\beta = .9$

For clinic A our posterior odds of disease in this patient is:

$$\begin{aligned} O_{post} &= O_{prior}\Omega_{X=1} \\ &= \frac{P}{Q}\frac{\alpha}{(1-\beta)} \\ &= \frac{.01}{.99}\frac{.875}{(1-.9)} \\ &= \frac{35}{396} \approx .0884, \end{aligned}$$

or a posterior probability of disease of $\frac{.0884}{(.0884+1)} = .0812$

Worked Example – Clinic B.

Now we move to clinic B, where the base rate is much higher at $P = .25$. α and β do not depend on the base rate so they remain $\alpha = .875$ and $\beta = .9$. For clinic B our posterior odds of disease in this patient is:

$$\begin{aligned} O_{post} &= O_{prior}\Omega_{X=1} \\ &= \frac{P}{Q}\frac{\alpha}{(1-\beta)} \\ &= \frac{.25}{.75}\frac{.875}{(1-.9)} \\ &= \frac{35}{12} \approx 2.917 \end{aligned}$$

or a posterior probability of disease of $\frac{2.917}{(2.917+1)} = .74$.

Worked Example – Two Tests

Let's say we'd like to be more confident in our diagnosis than 3:1 odds. If we have another test Z *and the test is independent from the previous test* it's a simple matter to use it to update our current best guess.

- Assume $\alpha_Z = .7$ and $\beta_Z = .95$
- Our old test we refer to as test X

$$\begin{aligned} O_{post} &= O_{prior}\Omega_{X=x_i}\Omega_{Z=z_i} \\ &= \frac{P}{Q} \times \begin{cases} \frac{1-\alpha_X}{\beta_X}\frac{1-\alpha_Z}{\beta_Z} & \text{if } X=0, Z=0; \\ \frac{1-\alpha_X}{\beta_X}\frac{\alpha_Z}{1-\beta_Z} & \text{if } X=0, Z=1; \\ \frac{\alpha_X}{1-\beta_X}\frac{1-\alpha_Z}{\beta_Z} & \text{if } X=1, Z=0; \text{ and} \\ \frac{\alpha_X}{1-\beta_X}\frac{\alpha_Z}{1-\beta_Z} & \text{if } X=1, Z=1; \end{cases} \end{aligned}$$

Two *independent* Tests

Let's say we'd like to be more confident in our diagnosis than 3:1 odds. If we have another test Z *and the test is independent from the previous test* it's a simple matter to use it to update our current best guess.

- Assume $\alpha_Z = .7$ and $\beta_Z = .95$
- Our old test we refer to as test X
- A positive test result on tests X and Z in clinic B results in:

$$\begin{aligned} O_{post} &= O_{prior}\Omega_{X=x_i}\Omega_{Z=z_i} \\ &= \frac{.25}{.75}\frac{.875}{1-.9}\frac{.7}{1-.95} \\ &= 40.833, \end{aligned}$$

a posterior probability of $\frac{40.833}{1+40.833} = .98$.

Two *independent* Tests

All possible outcomes for this patient in clinic B

$$O_{post} = O_{prior}\Omega_{X=x_i}\Omega_{Z=z_i}$$

$$= \frac{.25}{.75} \times \begin{cases} \frac{1-.875}{.9}\frac{1-.7}{.95} & = .002, \text{ when } X = 0, Z = 0 \\ \frac{1-.875}{.9}\frac{.7}{1-.95} & = .13, \text{ when } X = 0, Z = 1 \\ \frac{.875}{1-.9}\frac{1-.7}{.95} & = .77, \text{ when } X = 1, Z = 0 \\ \frac{.875}{1-.9}\frac{.7}{1-.95} & = 41, \text{ when } X = 1, Z = 1 \end{cases}$$

Total Relevant Evidence

- More ‘information’ is not necessarily better, especially when you’re a human judge.
- Independent and valid evidence (tests) is ideal. This is also rare.
- Combining information across multiple correlated tests is tricky and depends on the strength of correlation, which could be different in the cases versus controls.
- This issue will come up later when we discuss *incremental validity*.

Cutting Scores and (Quasi-) Continuous Tests

Show program. Note that cutting score affects sensitivity & specificity and thereby hit rate. Optimal cutting score depends on the test score distributions (separation, size, shape).

Betting the Base Rate

In some instances the base rate is so low, the tests are so weak and poorly calibrated, that it is more efficient to "bet the base rate" than administer the test.

Table: Test performance under varying base rates and cutting scores

BR	Scale	Cut Score	PPV	NPV	α	β	HR	Betting BR HR
.45	*F*	≥ 90	.72	1.0	1.0	.68	.82	.55
.45	*F*	≥ 100	.80	.97	.97	.80	.88	.55
.16	*F*	≥ 90	.38	1.0	1.0	.68	.74	.84
.16	*F*	≥ 100	.50	1.0	1.0	.81	.84	.84

MMPI F scale (Arbisi & Ben-Porath, Psych Assessment, 1998)

Simple Decision-Theoretic Analysis

- So far our goal has been to maximize correct classifications of disorder. If $O_{post} > 1$, diagnose disease. If $O_{post} < 1$ do not.
- We have ignored the costs of making different decisions, which may affect how cautious or aggressive we are in making diagnoses.

Simple Decision-Theoretic Analysis: Classic Example

Tomorrow morning I arise and prepare to go out. I have to decide whether to take my umbrella. If it rains and I have my umbrella, I shall remain dry; otherwise I'll get wet, which is disagreeable. On the other hand, if I carry my umbrella and it doesn't rain, I'm burdened with the clumsy thing, which is also unpleasant. The matrix of disutilities is:

		Weather	
		Dry	Rain
Umbrella	Don't Carry	5	50
	Carry	20	3

Correct for the minimum disutility per column to obtain:

		Weather	
		Dry	Rain
Umbrella	Don't Carry	0	47
	Carry	15	0

Simple Decision-Theoretic Analysis: Incorporating Base Rates

Now assume we know there's a 30% chance of rain tomorrow, we can revise our expected *disutilities*:

		Weather: Dry	Weather: Rain
Umbrella	Don't Carry	$(1 - .3) \times 0 = 0$	$.3 \times 47 = 14.1$
	Carry	$(1 - .3) \times 15 = 10.5$	$.3 \times 0 = 0$

If you were a rational actor, what would you do?

Complicated Decision-Theoretic Analysis

Now we have a patient expressing increased suicidal ideation, who has a plan, but denies intention to suicide. Grove states the base rate for suicide among such individuals is $\approx 1\%$. Let's make up some disutilities.

		Will Suicide	
		No	Yes
Hospitalize	No	0	225
	Yes	100	0

The expected disutility of not hospitalizing such patients is:

$$.99 \times 0 + .01 \times 225 = 2.25$$

The expected disutility of hospitalizing them is:

$$.99 \times 100 + .01 \times 0 = 99$$

What would a rational actor do?

Complicated Decision-Theoretic Analysis

What if the disutility matrix was different?

		Will Suicide	
		No	Yes
Hospitalize	No	0	1000
	Yes	1	0

The expected disutility of not hospitalizing such patients is:

$$.99 \times 0 + .01 \times 1000 = 10$$

The expected disutility of hospitalizing them is:

$$.99 \times 1 + .01 \times 0 = .99$$

Under this scenario it "pays" to hospitalize such patients.

Does this seem appealing?

- How difficult is it to do this?
- Who are the stakeholders, how do you measure their disutilities, how do you weight them?

Does this seem appealing?

- How difficult is it to do this?
- Who are the stakeholders, how do you measure their disutilities, how do you weight them?
- Is difficulty a reason to ignore these issues?