

T.8 · Conditioning, Bayesian inference and sampling error

GRE · GRE Subject Test · GRE Mathematics

Scope and prerequisites

Undergraduate GRE preparation. Local objectives within the reviewed ETS scope; this is original teaching, not an official test or score predictor.

Prerequisites: Probability tables, expectation, variance and independent sampling.

- Separate conditional, joint and independent-event probabilities
- Calculate posterior probabilities with a complete base-rate table
- Use variance and sample size to determine a sample mean standard error

posterior probability : A probability conditional on the observed evidence after accounting for prior proportions.

standard error : The standard deviation of a statistic across repeated samples.

Choose and justify a method

For $P(B)>0$, $P(A|B)=P(A\cap B)/P(B)$: restrict the population to B before computing the fraction in A. Independence means $P(A\cap B)=P(A)P(B)$, equivalently $P(A|B)=P(A)$ when the denominator is nonzero. Mutually exclusive events with positive probabilities cannot be independent, because their joint probability is zero. Sampling without replacement usually changes later probabilities; a fixed denominator for successive draws would describe a different experiment.

Build joint probabilities by multiplying a prior category proportion by the relevant conditional probability. If D is the condition and + a positive result, $P(D\cap+)=P(D)P(+|D)$. The total positive probability is $P(D)P(+|D)+P(\text{not } D)P(+|\text{not } D)$, because these are disjoint and cover every positive result. Bayes divides the first joint probability by this total. Sensitivity is $P(+|D)$, while specificity is $P(-|\text{not } D)$; the false-positive rate is one minus specificity.

For a hypothetical test with prevalence 2%, sensitivity 90% and specificity 95%, use a population of 10,000 for clarity. Of 200 people with the condition, 180 test positive. Of 9,800 without it, 490 test positive. Therefore 180 of 670 positive results have the condition: posterior 18/67, about 26.9%. Reversing the conditional would give 90%, answering a different question. This is a mathematical model, not a recommendation about clinical decisions.

For independent identically distributed observations with finite variance σ^2 , the sample mean has expectation μ and variance σ^2/n ; its standard error is $\sigma/\sqrt{n}$. This is spread of repeated sample means, not the spread of individual observations. Quadru-

pling n halves the standard error, rather than quartering it. Normal population data give an exactly normal mean; otherwise a central-limit approximation needs adequate conditions and sample size. Correlation invalidates the simple independent variance calculation.

Worked reasoning

Let $P(A)=0.4$, $P(B)=0.5$ and $P(A\cap B)=0.3$. Then $P(A|B)=0.6$, so A and B are not independent, since $0.3\neq 0.4 \cdot 0.5$. If a variable has standard deviation 12, an independent sample of size 36 has mean standard error 2. Increasing the sample size to 144 gives 1; the individual-observation standard deviation stays 12.

Condition on positives: 80 / 125, not 80 / 100

Expected group	Positive	Negative	Total
Defective	80	20	100
Not defective	45	855	900
Total	125	875	1000

Original course illustration; its values belong to the worked example, not the later practice.

Conditions and counterexamples

Do not reverse $P(+|D)$ into $P(D|+)$. Include false positives in the posterior denominator. Standard deviation of individual observations and standard error of their mean have different sample-size behaviour.

Guided application

In a stated model, 10% of devices have a defect. A test is positive for 80% of defective and 5% of nondefective devices. Find the probability of a defect given a positive result using 1000 expected devices.

■ Worked solution

Expected counts are 100 defective and 900 nondefective. Positive counts are $100(0.8) = 80$ and $900(0.05) = 45$, respectively. The conditioned group contains 125 positives, hence

$$P(D \mid +) = \frac{80}{80 + 45} = \frac{16}{25} = 0.64.$$

Negative counts 20 and 855 check both row totals. The sensitivity 0.8 is $P(+ \mid D)$, not the requested reversed condition. These are illustrative model rates, not measured performance claims.

Independent transfer

Independent observations have variance 25. Find the standard error of their mean for $n=100$. If instead every distinct pair has covariance 1, find the mean's variance. Does quadrupling n necessarily halve the standard error in that second model?

■ Check after attempting

Independence gives $\text{Var}(\bar{X}) = 25/n$ and at 100 standard error $5/10 = 0.5$. With covariance one,

$$\text{Var}(\bar{X}) = \frac{25n + n(n-1)}{n^2} = 1 + \frac{24}{n}.$$

At 100 it is 1.24. The standard error tends to one as n grows, rather than to zero; quadrupling n need not halve it. Covariance terms may be dropped only under justified uncorrelatedness, with independence a sufficient condition.