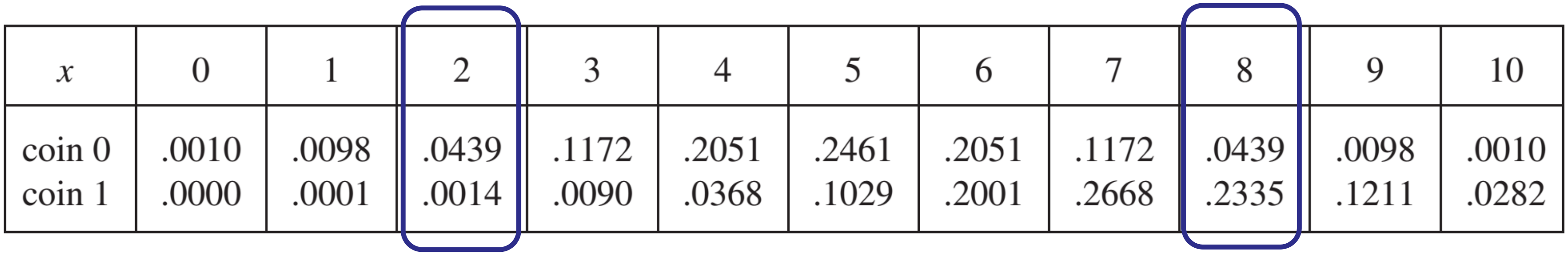
Testing Hypotheses and Assessing Goodness of Fit

Starting with an example – Coins again

I have two coins, coin 0 has probability of heads =0.5, coin 1 has 0.7. I choose one of the coins, toss it 10 times and tell you # of heads, but do not tell you whether it was coin 0 or coin 1.

On the basis of the number of heads, your task is to decide which coin it was. How should your decision rule be?

X= # of heads



x	0	1	2	3	4	5	6	7	8	9	10
coin 0	.0010	.0098	.0439	.1172	.2051	.2461	.2051	.1172	.0439	.0098	.0010
coin 1	.0000	.0001	.0014	.0090	.0368	.1029	.2001	.2668	.2335	.1211	.0282

The **likelihood ratio** —
We observed 2 heads, $P_0(2)/P_1(2)$~30, coin 0 favored.
We observed 8 heads, $P_0(8)/P_1(8)$~ 0.2, coin 1 favored.

Starting with an example – Coins again

Specify 2 hypotheses and develop a **Bayesian** methodology:

H_0: Coin 0 was tossed H_1: Coin 1 was tossed

Prior probabilities: $P(H_0)$ and $P(H_1)$. Maybe $P(H_0) = P(H_1) = 1/2.$

Posterior probabilities

$$P(H_0|x) = \frac{P(H_0, x)}{P(x)} = \frac{P(x|H_0)P(H_0)}{P(x)}$$

… and the ratio

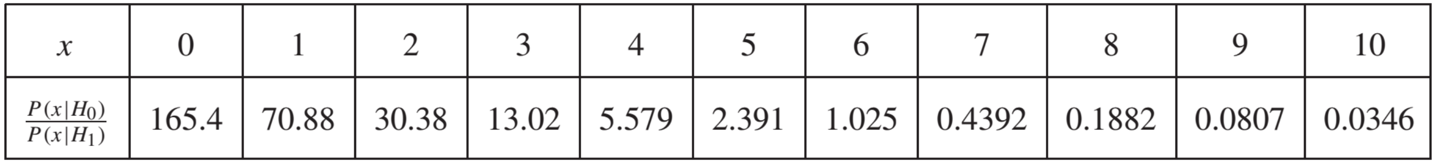
$$\frac{P(H_0|x)}{P(H_1|x)} = \frac{P(H_0)}{P(H_1)} \frac{P(x|H_0)}{P(x|H_1)}$$

Prior ratio times likelihood ratio

The evidence provided by the data is contained in the likelihood ratio.

Starting with an example – Coins again

Likelihood ratio



x	0	1	2	3	4	5	6	7	8	9	10
$\frac{P(x\|H_0)}{P(x\|H_1)}$	165.4	70.88	30.38	13.02	5.579	2.391	1.025	0.4392	0.1882	0.0807	0.0346

X decreases, H_0 increasingly favored; X increases, H_1 increasingly favored.

If $P(H_0) = P(H_1) = 1/2$ then:
0-6 heads, H_0 more probable; 7-10 heads, H_1 more probable.

Which one to choose? Larger posterior probability!

Choose H_0 if

$$\frac{P(H_0|x)}{P(H_1|x)} = \frac{P(H_0)}{P(H_1)} \frac{P(x|H_0)}{P(x|H_1)} > 1 \qquad \frac{P(x|H_0)}{P(x|H_1)} > c$$

Critical value c depends on priors. Your decision is based on likelihood ratio: accept H_0 if likelihood ratio > c, reject H_0 if likelihood ratio < c.

Starting with an example – Coins again

Consequences of a particular decision rule… suppose c=1.

0-6 heads, H_0 accepted; 7-10 heads, H_0 rejected.

Two possible errors: **Reject H_0 when it's true, accept H_0 when it's false**.

$$\begin{aligned} P(\text{reject } H_0|H_0) &= P(X > 6|H_0) \\ &= 0.18 \end{aligned}$$

$$\begin{aligned} P(\text{accept } H_0|H_1) &= P(X \leq 6|H_1) \\ &= 0.35 \end{aligned}$$

Likelihood ratio

1

x	0	1	2	3	4	5	6	7	8	9	10
$\frac{P(x\|H_0)}{P(x\|H_1)}$	165.4	70.88	30.38	13.02	5.579	2.391	1.025	0.4392	0.1882	0.0807	0.0346

Starting with an example – Coins again

Consequences of a particular decision rule… suppose c=1.
0-6 heads, H_0 accepted; 7-10 heads, H_0 rejected.

Two possible errors: **Reject H_0 when it's true, accept H_0 when it's false**.

$$P(\text{reject } H_0 | H_0) = P(X > 6 | H_0) = 0.18 \qquad P(\text{accept } H_0 | H_1) = P(X \leq 6 | H_1) = 0.35$$

Now say c = 0.1, $P(H_0)/P(H_1) = 10$:
0-8 heads, H_0 accepted; 9-10 heads, H_0 rejected – *more extreme rejection evidence*

$$P(\text{reject } H_0 | H_0) = P(X > 8 | H_0) = 0.01 \qquad P(\text{accept } H_0 | H_1) = P(X \leq 8 | H_1) = 0.85$$

***c* controls the *tradeoff* between the probabilities of the two types of errors.**

Likelihood ratio

1

0.1

x	0	1	2	3	4	5	6	7	8	9	10
$\frac{P(x\|H_0)}{P(x\|H_1)}$	165.4	70.88	30.38	13.02	5.579	2.391	1.025	0.4392	0.1882	0.0807	0.0346

The Neyman-Pearson paradigm

Classically, Neyman & Pearson formulated hypothesis testing by casting it as a decision problem and making the 2 types of errors central (no prior needed)

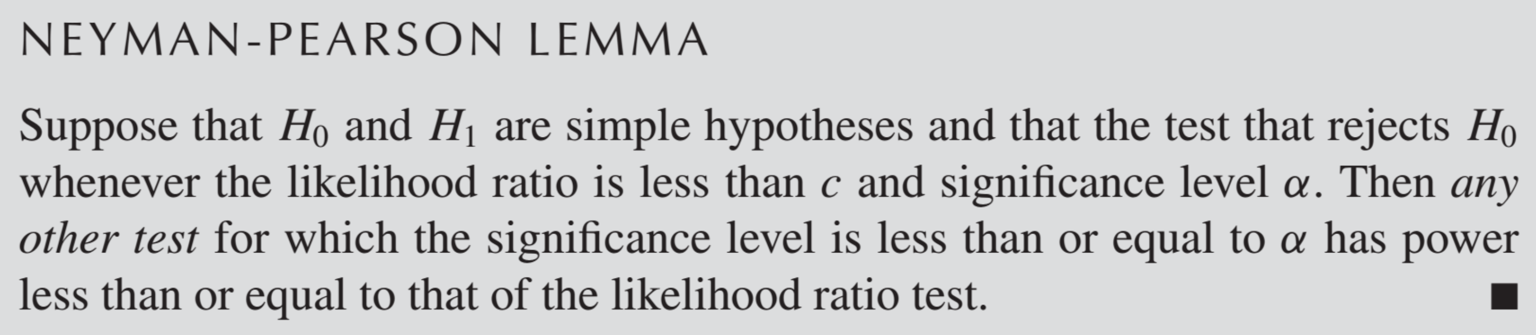
An asymmetry is introduced:
null hypothesis H_0 vs. **alternative hypothesis H_1/H_A**（原假设，备择假设）.

Standard terminology:

- Rejecting H_0 when it is true is called a **type I error.** “错杀”
- Probability of a type I error = the **significance level** of the test, denoted by α.
- Accepting H_0 when it is false is called a **type II error** and its probability is β. “错放”
- Probability that H_0 is rejected when it is false is the **power**（势） of the test = $1-\beta$.
- The likelihood ratio or # of heads above, is the **test statistic**（检验统计量）.
- The set of values of the test statistic leading to rejection of H_0 is the **rejection region,** the set of values leading to acceptance is the **acceptance region**（接受/拒绝域）.
- The probability distribution of the test statistic when H_0 is true is the **null distribution**（原分布，零分布）.

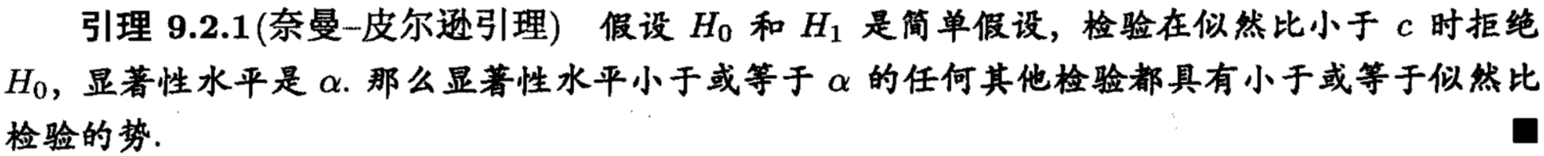
The Neyman-Pearson Lemma

A **simple hypothesis** completely specifies the probability distribution.
e.g. binomial (10, 0.5) or binomial (10, 0.7) in our example.



NEYMAN-PEARSON LEMMA

Suppose that H_0 and H_1 are simple hypotheses and that the test that rejects H_0 whenever the likelihood ratio is less than c and significance level α. Then *any other test* for which the significance level is less than or equal to α has power less than or equal to that of the likelihood ratio test. ■



引理 9.2.1(奈曼–皮尔逊引理) 假设 H_0 和 H_1 是简单假设，检验在似然比小于 c 时拒绝 H_0，显著性水平是 α. 那么显著性水平小于或等于 α 的任何其他检验都具有小于或等于似然比检验的势. ■

- Accepting H_0 when it is false is called a **type II error** and its probability is β.
- Probability that H_0 is rejected when it is false is the **power** of the test = $1-\beta$.

For simple hypotheses, basing the test on the likelihood ratio is optimal!

The Neyman-Pearson Paradigm

Let $X_1, \ldots, X_n$ be a random sample from a normal distribution having known variance σ^2. Consider two simple hypotheses:

$$H_0\colon \mu = \mu_0 \qquad \mu_0 - \mu_1 < 0$$
$$H_A\colon \mu = \mu_1$$

Neyman-Pearson Lemma: among all tests with significance level α, the test that rejects for small values of the likelihood ratio is most powerful.

$$\frac{f_0(\mathbf{X})}{f_1(\mathbf{X})} = \frac{\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_0)^2\right]}{\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_1)^2\right]}$$

test statistic

$$2n\overline{X}(\mu_0 - \mu_1) + n\mu_1^2 - n\mu_0^2$$

is a function of $\overline{X}$ and is small when $\overline{X}$ is large.

Rejection: $\overline{X} > x_0$ for some x_0, x_0 is chosen so that $P(\overline{X} > x_0) = \alpha$ if H_0 is true.

- Null distribution of $\overline{X}$ is Gaussian with mean=μ_0, variance=σ^2/n.

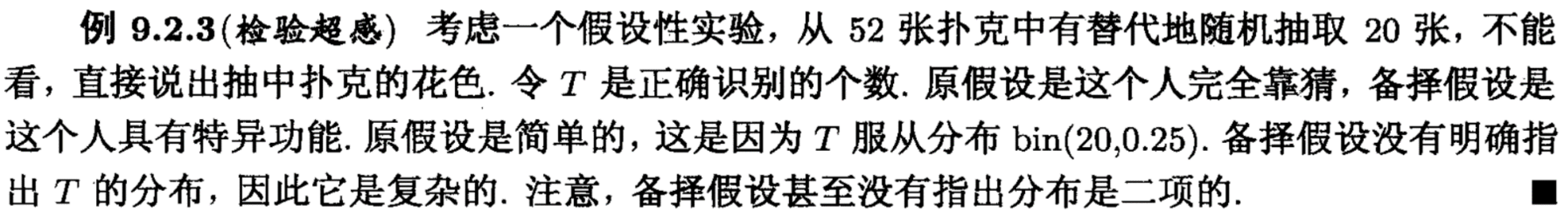
$$P(\overline{X} > x_0) = P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{x_0 - \mu_0}{\sigma/\sqrt{n}}\right) \qquad \frac{x_0 - \mu_0}{\sigma/\sqrt{n}} = z(\alpha)$$

The Neyman-Pearson Paradigm

Unfortunately, N-P Lemma is of little direct utility in most practical problems, the case of testing a simple null hypothesis vs. a simple alternative is rarely encountered.

A **composite hypothesis** does not completely specify the probability distribution.

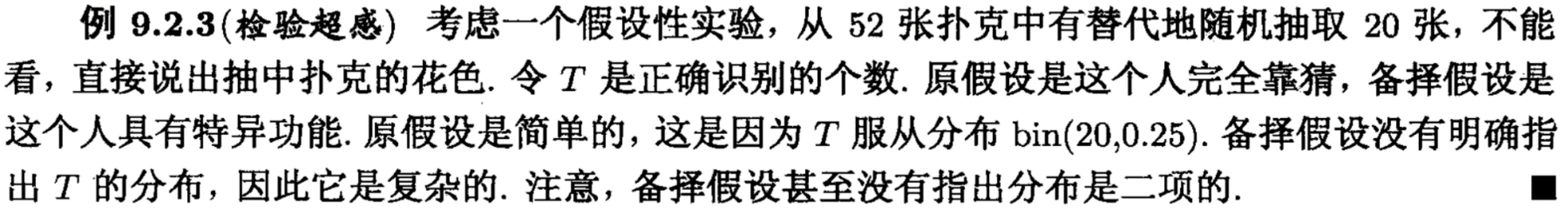
Example. Goodness-of-fit test
$X_1, \ldots, X_n$ is a sample from a discrete probability distribution.
H_0 : distribution is Poisson, with some <u>unspecified</u> mean (*composite hypothesis*)
H_1 : distribution is NOT Poisson (*composite hypothesis*)



例 9.2.3(检验超感)　考虑一个假设性实验，从 52 张扑克中有替代地随机抽取 20 张，不能看，直接说出抽中扑克的花色. 令 T 是正确识别的个数. 原假设是这个人完全靠猜，备择假设是这个人具有特异功能. 原假设是简单的，这是因为 T 服从分布 bin(20,0.25). 备择假设没有明确指出 T 的分布，因此它是复杂的. 注意，备择假设甚至没有指出分布是二项的. ■

Specification of the significance and the concept of a *p*-value

Neyman-Pearson approach has the strength that only the distribution under H_0 is needed in order to construct a test.



例 9.2.3(检验超感) 考虑一个假设性实验，从 52 张扑克中有替代地随机抽取 20 张，不能看，直接说出抽中扑克的花色. 令 T 是正确识别的个数. 原假设是这个人完全靠猜，备择假设是这个人具有特异功能. 原假设是简单的，这是因为 T 服从分布 bin(20,0.25). 备择假设没有明确指出 T 的分布，因此它是复杂的. 注意，备择假设甚至没有指出分布是二项的. ■

Large T supports H_1, so rejection region has the form $\{T \geq t_0\}$, t_0 is chosen so that $P(T \geq t_0 | H_0)$ = the desired significance level of the test α.

No need to specify the probability distribution under H_1, but only notice that if H_1 is true, the subject would correctly identify more suits than if purely guessing.

In comparison, *a fully Bayesian treatment would have to specify the distribution under the alternative as well as prior probabilities*.

Specification of the significance and the concept of a *p*-value

Criticism of the paradigm:

- How to choose α? The theory requires the specification of α, in advance of analyzing the data, but gives no "how-to" guidance.
- In practice it is almost always the case that the choice of α is essentially arbitrary, but is heavily influenced by custom. Small values, such as 0.01 and 0.05, are commonly used.
- It is built on the assumption that one must either reject or not reject a hypothesis, when typically no such decision is actually required. The theory is thus often applied in a hypothetical manner.

Example: the subject above guessed the suit correctly 9 times. Since $P(T \geq 9|H_0) =$ 0.041, H_0 would have been "rejected" at the significance level $\alpha = .05$.

A ***p*-value** is defined to be the smallest significance level at which H_0 would be rejected. (*p*-value = 0.041 here.)

Ronald Fisher: *p*-value is the probability under H_0 of a result as or more extreme than that actually observed (e.g. chance of getting at least 9 correct by guessing).

Notes on the null hypothesis

There is an asymmetry in the Neyman-Pearson paradigm between the null and alternative hypotheses.

- It is conventional to choose the **simpler** of two hypotheses as the null, e.g. H_0: distribution is Poisson, H_1: it is not Poisson.

- The consequences of incorrectly rejecting one hypothesis may be graver than the other -- the former should be chosen as H_0, because the probability of falsely rejecting it could be controlled by choosing α (e.g. screening new drugs).

- In scientific investigations, the null hypothesis is often a simple explanation that must be discredited in order to demonstrate the presence of some physical phenomenon or effect (科学精神是质疑，如超能力检验的例子：）The validity of the null hypothesis (purely guessing) would not be cast in doubt *unless* the results would be extremely unlikely under the null.

Generalized Likelihood Ratio Tests

Generalized likelihood ratio tests

The likelihood ratio test is optimal for simple vs. simple hypotheses.
Generalized likelihood ratio tests are for use when hypotheses are not simple.

They are not generally optimal, but are typically non-optimal in situations where no optimal test exists, and they usually perform reasonably well.

Wide utility, plays the same role in testing as MLE's do in estimation.

Generalized likelihood ratio tests

The likelihood ratio test is optimal for simple vs. simple hypotheses.
Generalized likelihood ratio tests are for use when hypotheses are not simple.

They are not generally optimal, but are typically non-optimal in situations where no optimal test exists, and they usually perform reasonably well.

Wide utility, plays the same role in testing as MLE's do in estimation.

- Suppose that the observations $\boldsymbol{X} = (X_1, \ldots, X_n)$ have a joint distribution $f(\mathbf{x}|\theta)$.
- H_0 may specify that $\theta \in \omega_0$, where ω_0 is a subset of the set of all possible θ values
- H_1 may specify that $\theta \in \omega_1$, where ω_1 is disjoint from ω_0.
- Let $\Omega = \omega_0 \cup \omega_1$.
- A plausible measure of the relative tenability of the hypotheses is the ratio of their likelihoods.
- Generalized likelihood ratio: discredit H_0 if small

$$\Lambda^* = \frac{\max\limits_{\theta \in \omega_0}[\text{lik}(\theta)]}{\max\limits_{\theta \in \omega_1}[\text{lik}(\theta)]}$$

Generalized likelihood ratio tests

The likelihood ratio test is optimal for simple vs. simple hypotheses.
Generalized likelihood ratio tests are for use when hypotheses are not simple.

They are not generally optimal, but are typically non-optimal in situations where no optimal test exists, and they usually perform reasonably well.

Wide utility, plays the same role in testing as MLE's do in estimation.

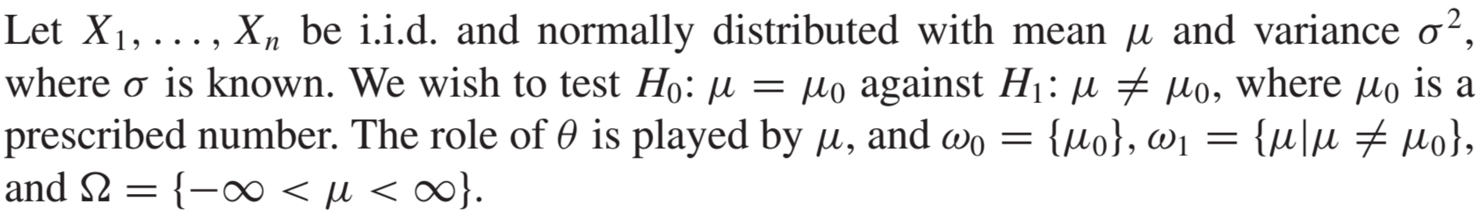
- Suppose that the observations $\boldsymbol{X} = (X_1, \ldots, X_n)$ have a joint distribution $f(\mathbf{x}|\theta)$.
- H_0 may specify that $\theta \in \omega_0$, where ω_0 is a subset of the set of all possible θ values
- H_1 may specify that $\theta \in \omega_1$, where ω_1 is disjoint from ω_0.
- Let $\Omega = \omega_0 \cup \omega_1$.
- A plausible measure of the relative tenability of the hypotheses is the ratio of their likelihoods.
- Generalized likelihood ratio: discredit H_0 if small $$\Lambda^* = \frac{\max\limits_{\theta \in \omega_0}[\text{lik}(\theta)]}{\max\limits_{\theta \in \omega_1}[\text{lik}(\theta)]}$$
- In practical, use, instead, $\Lambda = \dfrac{\max\limits_{\theta \in \omega_0}[\text{lik}(\theta)]}{\max\limits_{\theta \in \Omega}[\text{lik}(\theta)]}$ and note $\Lambda = \min(\Lambda^*, 1)$
- Rejection region: small values.

  $\Lambda \le \lambda_0$. The threshhold λ_0 is chosen so that $P(\Lambda \le \lambda_0 | H_0) = \alpha$,

  desired significance level of the test.

Generalized likelihood ratio tests

Example. Testing a normal mean.



Let $X_1, \ldots, X_n$ be i.i.d. and normally distributed with mean μ and variance σ^2, where σ is known. We wish to test H_0: $\mu = \mu_0$ against H_1: $\mu \neq \mu_0$, where μ_0 is a prescribed number. The role of θ is played by μ, and $\omega_0 = \{\mu_0\}$, $\omega_1 = \{\mu | \mu \neq \mu_0\}$, and $\Omega = \{-\infty < \mu < \infty\}$.

Numerator of likelihood ratio: $\dfrac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_0)^2}$

Denominator: ???

Generalized likelihood ratio tests

Example. Testing a normal mean.

Let $X_1, \ldots, X_n$ be i.i.d. and normally distributed with mean μ and variance σ^2, where σ is known. We wish to test H_0: $\mu = \mu_0$ against H_1: $\mu \neq \mu_0$, where μ_0 is a prescribed number. The role of θ is played by μ, and $\omega_0 = \{\mu_0\}$, $\omega_1 = \{\mu | \mu \neq \mu_0\}$, and $\Omega = \{-\infty < \mu < \infty\}$.

Numerator of likelihood ratio: $\dfrac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu_0)^2}$

Denominator:

(MLE: $\mu = \overline{X}$) $\mu \in \Omega$ $\dfrac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2}$

Likelihood ratio statistic is

$$\Lambda = \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (X_i - \mu_0)^2 - \sum_{i=1}^n (X_i - \overline{X})^2\right]\right)$$

Rejection region is small Λ values, or large

$$-2\log\Lambda = \frac{1}{\sigma^2}\left(\sum_{i=1}^n (X_i - \mu_0)^2 - \sum_{i=1}^n (X_i - \overline{X})^2\right)$$

Generalized likelihood ratio tests

Rejection region is small Λ values, or large

$$-2\log\Lambda = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(X_i-\mu_0)^2-\sum_{i=1}^{n}(X_i-\overline{X})^2\right)$$

$$= n(\overline{X}-\mu_0)^2/\sigma^2$$

Q: What is its distribution?

Generalized likelihood ratio tests

Rejection region is small Λ values, or large

$$-2\log\Lambda = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(X_i-\mu_0)^2 - \sum_{i=1}^{n}(X_i-\overline{X})^2\right)$$

$$= n(\overline{X}-\mu_0)^2/\sigma^2$$

$$\boxed{\overline{X} \sim N(\mu_0, \sigma^2/n)} \qquad -2\log\Lambda \sim \chi_1^2$$

Construction of a rejection region for any significance level α: the test rejects when

$$\frac{n}{\sigma^2}(\overline{X}-\mu_0)^2 > \chi_1^2(\alpha)$$

Rejection region is

$$|\overline{X}-\mu_0| \geq \frac{\sigma}{\sqrt{n}}z(\alpha/2)$$

Generalized likelihood ratio tests

Rejection region is small Λ values, or large

$$-2\log\Lambda = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(X_i-\mu_0)^2 - \sum_{i=1}^{n}(X_i-\overline{X})^2\right)$$

$$= n(\overline{X}-\mu_0)^2/\sigma^2$$

$$\boxed{\overline{X} \sim N(\mu_0, \sigma^2/n)} \qquad -2\log\Lambda \sim \chi_1^2$$

Construction of a rejection region for any significance level α: the test rejects when

$$\frac{n}{\sigma^2}(\overline{X}-\mu_0)^2 > \chi_1^2(\alpha)$$

Rejection region is $|\overline{X}-\mu_0| \geq \frac{\sigma}{\sqrt{n}}z(\alpha/2)$

Wait! This is actually very obvious!

H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$

The test does not reject when μ_0 lies in a 100(1−α)% confidence interval for μ.

Generalized likelihood ratio tests

So far so good, but…

In order for the likelihood ratio test to have the significance level α, λ_0 must be chosen so that $P(\Lambda \leq \lambda_0)=\alpha$ if H_0 is true.

- If the sampling distribution of Λ under H_0 is known, we can determine λ_0.
- Generally, the sampling distribution is not of a simple form, what to do?

χ_1^2.

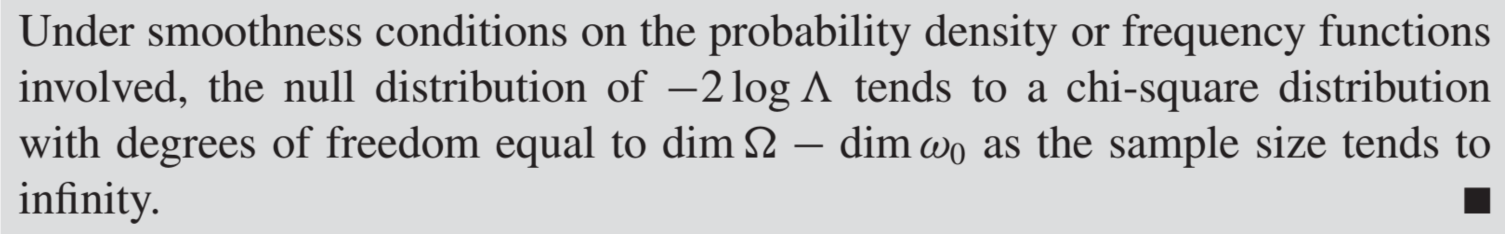
Generalized likelihood ratio tests

So far so good, but…

In order for the likelihood ratio test to have the significance level α, λ_0 must be chosen so that $P(\Lambda \leq \lambda_0)=\alpha$ if H_0 is true.

- If the sampling distribution of Λ under H_0 is known, we can determine λ_0.
- Generally, the sampling distribution is not of a simple form, what to do?

A theorem as the basis for an approximation to the null distribution:



Under smoothness conditions on the probability density or frequency functions involved, the null distribution of $-2\log\Lambda$ tends to a chi-square distribution with degrees of freedom equal to $\dim\Omega - \dim\omega_0$ as the sample size tends to infinity. ■

Dim Ω and dim ω_0 are # of free parameters under Ω and ω_0, respectively.

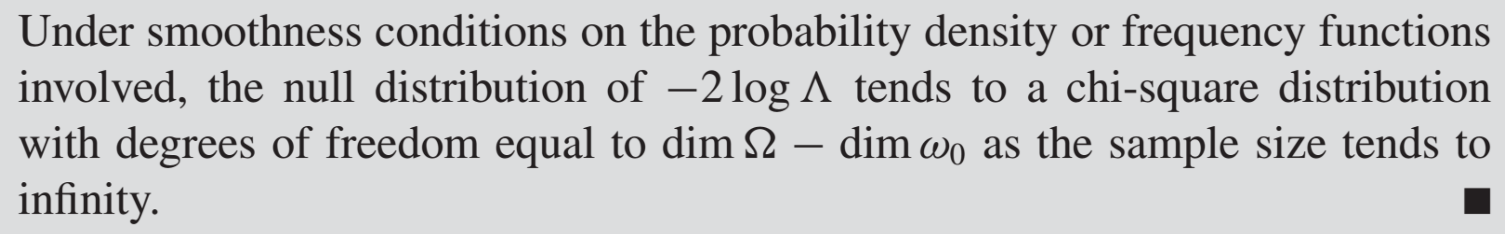
Generalized likelihood ratio tests

So far so good, but…

In order for the likelihood ratio test to have the significance level α, λ_0 must be chosen so that $P(\Lambda \leq \lambda_0)=\alpha$ if H_0 is true.

- If the sampling distribution of Λ under H_0 is known, we can determine λ_0.
- Generally, the sampling distribution is not of a simple form, what to do?

A theorem as the basis for an approximation to the null distribution:



Under smoothness conditions on the probability density or frequency functions involved, the null distribution of $-2\log\Lambda$ tends to a chi-square distribution with degrees of freedom equal to $\dim\Omega - \dim\omega_0$ as the sample size tends to infinity. ■

Dim Ω and dim ω_0 are # of free parameters under Ω and ω_0, respectively.

The above Gaussian example:

- H_0 completely specifies μ and σ^2, dim $\omega_0 = 0$
- Under Ω, σ is fixed, μ is free, dim $\Omega = 1$
- Null distribution of -2 log Λ is approximately (exactly!) χ_1^2.

Generalized likelihood Ratio Tests for the Multinomial Distribution

H_0 specifies that $p = p(\theta)$, $\theta \in \omega_0$, θ is a parameter (may be unknown);
H_1 cell probabilities are free except that they are nonnegative and sum to 1.
# of cells is m, Ω is the set consisting of m nonnegative numbers summing to 1.

Likelihood ratio's numerator is maximized when the MLE $\hat{\theta}$ is in place of θ:

$$\max_{p \in \omega_0} \left(\frac{n!}{x_1! \cdots x_m!} \right) p_1(\theta)^{x_1} \cdots p_m(\theta)^{x_m} \longrightarrow \boxed{p_i(\hat{\theta})}$$

Probabilities are unrestricted under Ω, denominator is maximized by the unrestricted MLE's,

$$\hat{p}_i = \frac{x_i}{n}$$

The likelihood ratio is, therefore,

$$\Lambda = \frac{\dfrac{n!}{x_1! \cdots x_m!} p_1(\hat{\theta})^{x_1} \cdots p_m(\hat{\theta})^{x_m}}{\dfrac{n!}{x_1! \cdots x_m!} \hat{p}_1^{x_1} \cdots \hat{p}_m^{x_m}} = \prod_{i=1}^{m} \left(\frac{p_i(\hat{\theta})}{\hat{p}_i} \right)^{x_i}$$

Also, since $x_i = n\hat{p}_i$,

$$-2 \log \Lambda = -2n \sum_{i=1}^{m} \hat{p}_i \log \left(\frac{p_i(\hat{\theta})}{\hat{p}_i} \right)$$

Generalized likelihood Ratio Tests for the Multinomial Distribution

$$-2\log\Lambda = -2n\sum_{i=1}^{m}\hat{p}_i \log\left(\frac{p_i(\hat{\theta})}{\hat{p}_i}\right)$$
$$= 2\sum_{i=1}^{m} O_i \log\left(\frac{O_i}{E_i}\right)$$

where $O_i = n\hat{p}_i$ and $E_i = np_i(\hat{\theta})$ denote the observed and expected counts

- Under Ω, cell probabilities are free but summing to 1, so dim $\Omega=m-1$
- Under H_0, probabilities $p_i(\hat{\theta})$ depend on a k-dimension parameter θ, dim $\omega_0=k$
- The large sample distribution of $-2\log\Lambda$ is chi-square with dof=$m-k-1$.

Pearson's chi-square statistic is commonly used to test for goodness of fit

$$\boxed{\chi^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i}}$$

$$X^2 = \sum_{i=1}^{m}\frac{[x_i - np_i(\hat{\theta})]^2}{np_i(\hat{\theta})}$$

Pearson's chi-square test

$$\boxed{\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}} \qquad X^2 = \sum_{i=1}^{m} \frac{[x_i - np_i(\hat{\theta})]^2}{np_i(\hat{\theta})}$$

Likelihood ratio

$$-2\log\Lambda = 2n \sum_{i=1}^{m} \hat{p}_i \log\left(\frac{\hat{p}_i}{p_i(\hat{\theta})}\right)$$

If H_0 is true, n is large, $\hat{p}_i \approx p_i(\hat{\theta})$. Taylor series expansion of the function about x_0 is

$$f(x) = x\log\left(\frac{x}{x_0}\right) \qquad f(x) = (x - x_0) + \frac{1}{2}(x - x_0)^2\frac{1}{x_0} + \cdots$$

So,

$$-2\log\Lambda \approx \underbrace{2n\sum_{i=1}^{m}[\hat{p}_i - p_i(\hat{\theta})]}_{\text{0, because } \Sigma p=1} + n\sum_{i=1}^{m}\frac{[\hat{p}_i - p_i(\hat{\theta})]^2}{p_i(\hat{\theta})}$$

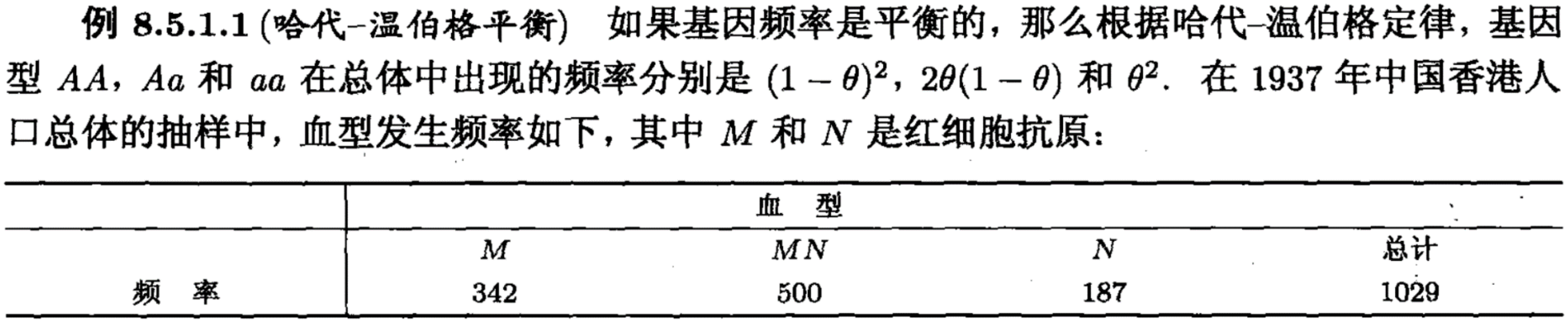
$$\boxed{\sum_{i=1}^{m}\frac{[x_i - np_i(\hat{\theta})]^2}{np_i(\hat{\theta})}}$$

Pearson's statistic & likelihood ratio are asymptotically equivalent under H_0.

But Pearson's chi-square test is easier in calculation and more widely used.

Pearson’s chi-square test: examples

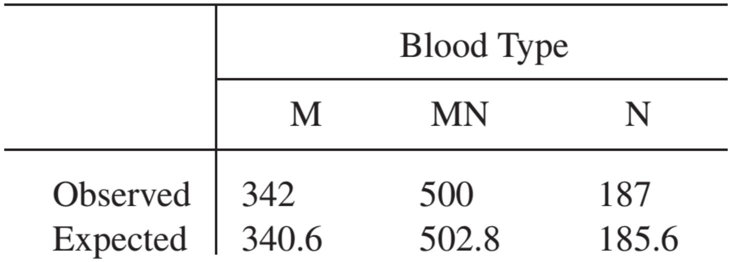
Hardy-Weinberg Equilibrium. Let’s test whether the model fits the data.



例 8.5.1.1 (哈代–温伯格平衡)　如果基因频率是平衡的，那么根据哈代–温伯格定律，基因型 AA，Aa 和 aa 在总体中出现的频率分别是 $(1-\theta)^2$，$2\theta(1-\theta)$ 和 θ^2. 在 1937 年中国香港人口总体的抽样中，血型发生频率如下，其中 M 和 N 是红细胞抗原：

	血型			
	M	MN	N	总计
频率	342	500	187	1029

MLE for θ gives $\hat{\theta} = .4247$, multiplying the resulting probabilities by sample size n=1029, calculate expected counts, compare with observations:



	Blood Type		
	M	MN	N
Observed	342	500	187
Expected	340.6	502.8	185.6

H_0: multinomial distribution is as specified by the H-W frequencies, unknown θ
H_1: multinomial distribution does not have probabilities of that specified form

Pearson's chi-square test: examples

Hardy-Weinberg Equilibrium. Let's test whether the model fits the data.

Significance level for the test, α（“错杀”概率）, set to be 0.05 for no reason.

Large sample: $-2 \log \Lambda \sim \chi^2$ (theorem), $-2 \log \Lambda \sim$ Pearson X^2, so Pearson $X^2 \sim \chi^2$

D.o.f. = dim Ω (3 cells − constraint of summing to 1) − dim ω_0 (1 parameter) = 1

The upper 5% of χ^2_1 is 3.84, so the test rejects if $X^2 > 3.84$.

Let's calculate X^2: $X^2 = \sum \frac{(O-E)^2}{E} = .0319$ So H_0 is not rejected.

Pearson’s chi-square test: examples

Hardy-Weinberg Equilibrium. Let’s test whether the model fits the data.

Significance level for the test, α（“错杀”概率）, set to be 0.05 for no reason.

Large sample: $-2 \log \Lambda \sim \chi^2$ (theorem), $-2 \log \Lambda \sim$ Pearson X^2, so Pearson $X^2 \sim \chi^2$

D.o.f. = dim Ω (3 cells − constraint of summing to 1) − dim ω_0 (1 parameter) = 1

The upper 5% of χ^2_1 is 3.84, so the test rejects if $X^2 > 3.84$.

Let’s calculate X^2: $$X^2 = \sum \frac{(O-E)^2}{E} = .0319$$ So H_0 is not rejected.

Again, no reason to choose α=0.05, and why should we have to make the decision?
Let’s use *p*-value instead! (smallest significance level at which H_0 would be rejected).

Probability that a chi-square random variable $\sim\chi^2_1$ is ≥ 0.0319 is 0.86=*p*-value:
if the model was correct, deviations this large or larger would occur 86% of the time.

Pearson’s chi-square test: examples

Hardy-Weinberg Equilibrium. Let’s test whether the model fits the data.

Significance level for the test, α（“错杀”概率）, set to be 0.05 for no reason.

Large sample: $-2 \log \Lambda \sim \chi^2$ (theorem), $-2 \log \Lambda \sim$ Pearson X^2, so Pearson $X^2 \sim \chi^2$

D.o.f. = dim Ω (3 cells − constraint of summing to 1) − dim ω_0 (1 parameter) = 1

The upper 5% of χ^2_1 is 3.84, so the test rejects if $X^2 > 3.84$.

Let’s calculate X^2: $X^2 = \sum \frac{(O-E)^2}{E} = .0319$ So H_0 is not rejected.

Again, no reason to choose α=0.05, and why should we have to make the decision? Let’s use *p*-value instead! (smallest significance level at which H_0 would be rejected).

Probability that a chi-square random variable $\sim\chi^2_1$ is ≥0.0319 is 0.86=*p*-value: if the model was correct, deviations this large or larger would occur 86% of the time.

- Likelihood ratio test statistic is the same: $-2 \log \Lambda = 2 \sum_{i=1}^{3} O_i \log \left(\frac{O_i}{E_i} \right) = .0319$

Pearson’s chi-square test: examples

Hardy-Weinberg Equilibrium. Let’s test whether the model fits the data.

Significance level for the test, α（“错杀”概率）, set to be 0.05 for no reason.

Large sample: $-2 \log \Lambda \sim \chi^2$ (theorem), $-2 \log \Lambda \sim$ Pearson X^2, so Pearson $X^2 \sim \chi^2$

D.o.f. = dim Ω (3 cells − constraint of summing to 1) − dim ω_0 (1 parameter) = 1

The upper 5% of $\chi^2{}_1$ is 3.84, so the test rejects if $X^2 > 3.84$.

Let’s calculate X^2: $X^2 = \sum \frac{(O-E)^2}{E} = .0319$ So H_0 is not rejected.

Again, no reason to choose α=0.05, and why should we have to make the decision?
Let’s use *p*-value instead! (smallest significance level at which H_0 would be rejected).

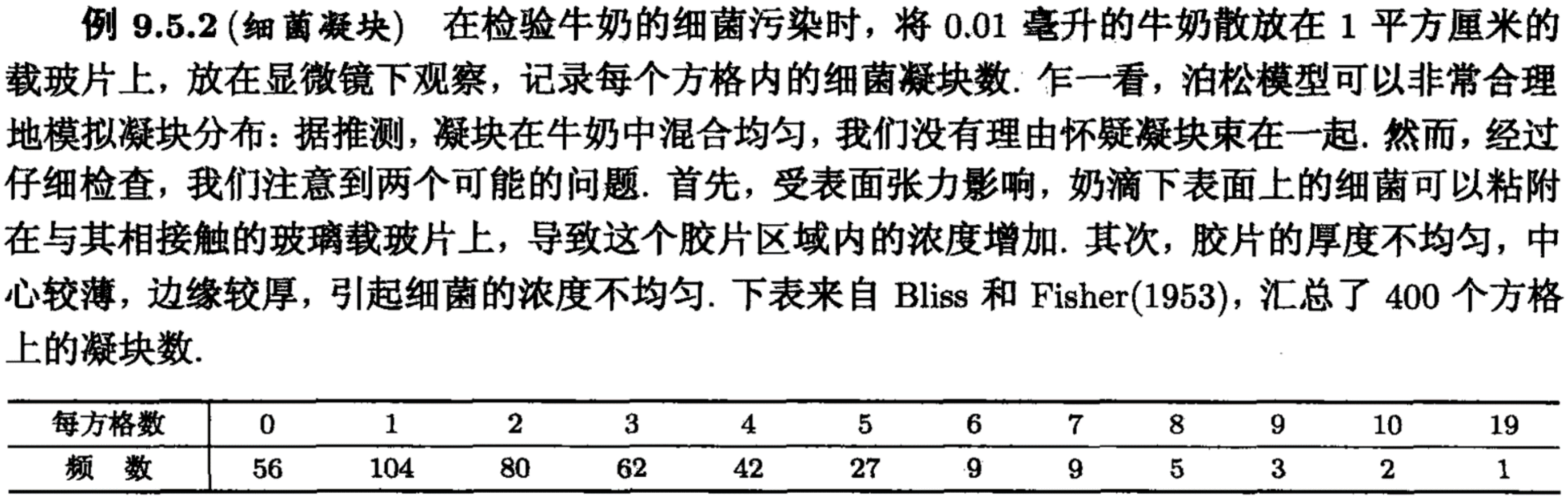
Probability that a chi-square random variable $\sim\chi^2{}_1$ is ≥0.0319 is 0.86=*p*-value:
if the model was correct, deviations this large or larger would occur 86% of the time.

- Likelihood ratio test statistic is the same: $-2 \log \Lambda = 2 \sum_{i=1}^{3} O_i \log \left(\frac{O_i}{E_i} \right) = .0319$

- Actual max likelihood ratio $\Lambda = \exp(-.0319/2) = .98.$ So H-W model is almost as likely as the most general possible model!

Pearson’s chi-square test: examples

Bacterial clumps. Let’s test whether Poisson fits the data.



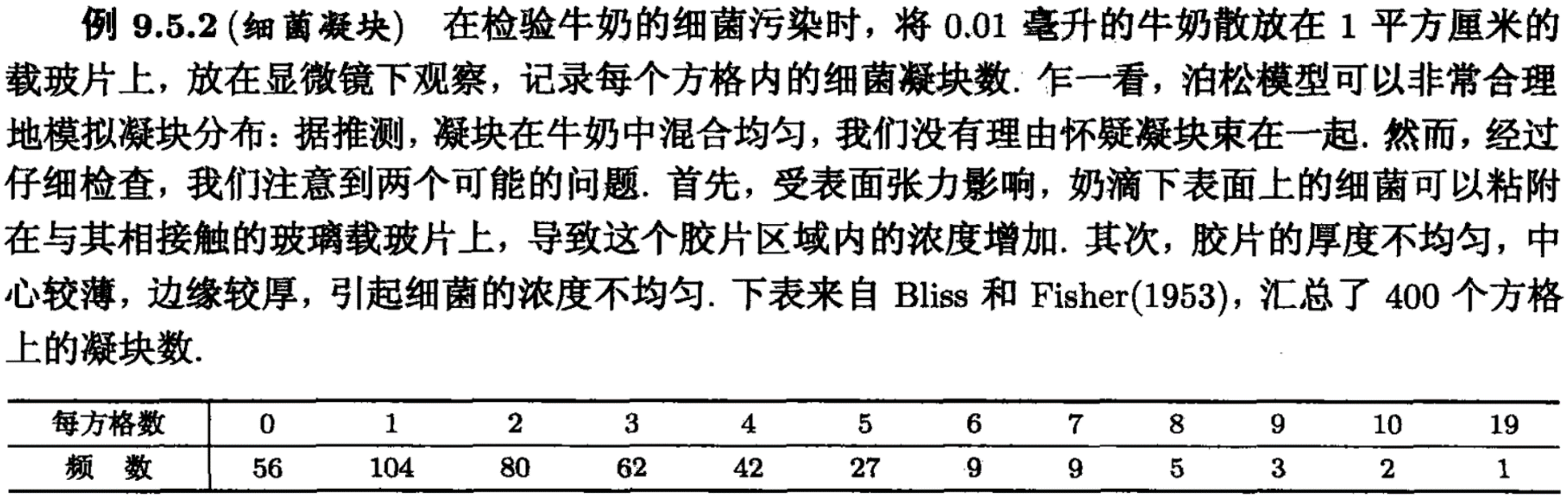
例 9.5.2 (**细菌凝块**)　在检验牛奶的细菌污染时，将 0.01 毫升的牛奶散放在 1 平方厘米的载玻片上，放在显微镜下观察，记录每个方格内的细菌凝块数. 乍一看，泊松模型可以非常合理地模拟凝块分布：据推测，凝块在牛奶中混合均匀，我们没有理由怀疑凝块束在一起. 然而，经过仔细检查，我们注意到两个可能的问题. 首先，受表面张力影响，奶滴下表面上的细菌可以粘附在与其相接触的玻璃载玻片上，导致这个胶片区域内的浓度增加. 其次，胶片的厚度不均匀，中心较薄，边缘较厚，引起细菌的浓度不均匀. 下表来自 Bliss 和 Fisher(1953)，汇总了 400 个方格上的凝块数.

每方格数	0	1	2	3	4	5	6	7	8	9	10	19
频　数	56	104	80	62	42	27	9	9	5	3	2	1

MLE gives us Poisson model with $\hat{\lambda} = \dfrac{0 \times 56 + 1 \times 104 + 2 \times 80 + \cdots + 19 \times 1}{400} = 2.44$

Pearson’s chi-square test: examples

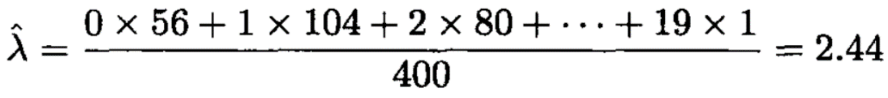
Bacterial clumps. Let’s test whether Poisson fits the data.



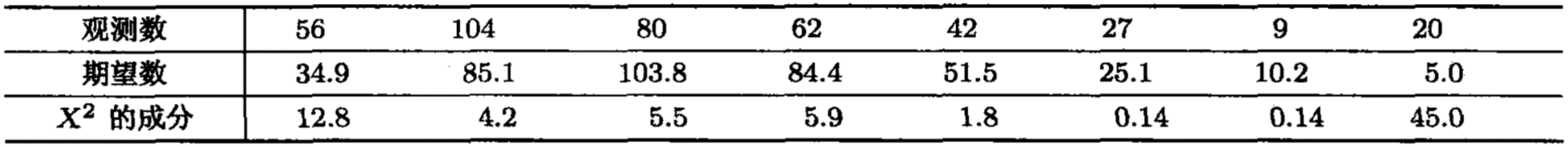
例 9.5.2 (细菌凝块) 在检验牛奶的细菌污染时，将 0.01 毫升的牛奶散放在 1 平方厘米的载玻片上，放在显微镜下观察，记录每个方格内的细菌凝块数. 乍一看，泊松模型可以非常合理地模拟凝块分布：据推测，凝块在牛奶中混合均匀，我们没有理由怀疑凝块束在一起. 然而，经过仔细检查，我们注意到两个可能的问题. 首先，受表面张力影响，奶滴下表面上的细菌可以粘附在与其相接触的玻璃载玻片上，导致这个胶片区域内的浓度增加. 其次，胶片的厚度不均匀，中心较薄，边缘较厚，引起细菌的浓度不均匀. 下表来自 Bliss 和 Fisher(1953)，汇总了 400 个方格上的凝块数.

每方格数	0	1	2	3	4	5	6	7	8	9	10	19
频　数	56	104	80	62	42	27	9	9	5	3	2	1

MLE gives us Poisson model with



$$\hat{\lambda} = \frac{0 \times 56 + 1 \times 104 + 2 \times 80 + \cdots + 19 \times 1}{400} = 2.44$$



观测数	56	104	80	62	42	27	9	20
期望数	34.9	85.1	103.8	84.4	51.5	25.1	10.2	5.0
X^2 的成分	12.8	4.2	5.5	5.9	1.8	0.14	0.14	45.0

H_0: Poisson, H_1: multinomial
Chi-square statistic is X^2=75.4, dof=dim Ω (8−1)−dim ω_0(1) = 6.
Null hypothesis is conclusively rejected (p-value<0.005).

Reason: too many small counts and large counts to be Poisson.