习题 7.4

1． 设 $X_1, \cdots, X_n$ 为来自 $b(1, p)$的样本，试求假设 H_0：$p = p_0$ vs H_1：$p \ne p_0$的似然比检验.

解：因样本联合密度函数为

$$p(x_1, \cdots, x_n; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} = p^{n\bar{x}}(1-p)^{n(1-\bar{x})},$$

则似然函数 $L(p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$，有 $\ln L(p) = n\bar{x}\ln p + n(1-\bar{x})\ln(1-p)$，

令 $\dfrac{d\ln L(p)}{dp} = \dfrac{n\bar{x}}{p} - \dfrac{n(1-\bar{x})}{1-p} = 0$，得 $p = \bar{x}$，即最大似然估计 $\hat{p} = \bar{X}$，

则 $\sup\limits_{p} L(p) = \bar{x}^{n\bar{x}}(1-\bar{x})^{n(1-\bar{x})}$，

当 $p = p_0$ 时，似然函数 $L(p_0) = p_0^{n\bar{x}}(1-p_0)^{n(1-\bar{x})}$，即 $\sup L(p_0) = L(p_0) = p_0^{n\bar{x}}(1-p_0)^{n(1-\bar{x})}$，

故似然比检验统计量为 $\Lambda(X_1, \cdots, X_n) = \dfrac{\sup\limits_{p} L(p)}{L(p_0)} = \left(\dfrac{\bar{X}}{p_0}\right)^{n\bar{X}}\left(\dfrac{1-\bar{X}}{1-p_0}\right)^{n(1-\bar{X})}$.

2． 设 $X_1, \cdots, X_n$ 为来自正态分布 $N(\mu, \sigma^2)$的样本，试求假设 H_0：$\sigma^2 = \sigma_0^2$ vs H_1：$\sigma^2 \ne \sigma_0^2$ 的似然比检验.

解：因样本联合密度函数为

$$p(x_1, \cdots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi)^{-\frac{n}{2}}\sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

则似然函数 $L(\mu, \sigma^2) = (2\pi)^{-\frac{n}{2}}\sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$，

有 $\ln L(\mu, \sigma^2) = -\dfrac{n}{2}\ln(2\pi) - \dfrac{n}{2}\ln\sigma^2 - \dfrac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)^2$，

令 $\begin{cases} \dfrac{\partial \ln L}{\partial \mu} = -\dfrac{1}{2\sigma^2}\sum\limits_{i=1}^{n} 2(x_i-\mu)\cdot(-1) = \dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu) = \dfrac{1}{\sigma^2}\left(\sum\limits_{i=1}^{n} x_i - n\mu\right) = 0; \\ \dfrac{\partial \ln L}{\partial \sigma^2} = -\dfrac{n}{2}\cdot\dfrac{1}{\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\sum\limits_{i=1}^{n}(x_i-\mu)^2 = 0. \end{cases}$

得 $\mu = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i = \bar{x}$，$\sigma^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\mu)^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$，

则 $\sup\limits_{\mu,\sigma^2} L(\mu, \sigma^2) = (2\pi)^{-\frac{n}{2}}\left[\dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\right]^{-\frac{n}{2}} e^{-\frac{n}{2}} = (2\pi)^{-\frac{n}{2}}\left[\dfrac{(n-1)s^2}{n}\right]^{-\frac{n}{2}} e^{-\frac{n}{2}}$，

当 $\sigma^2 = \sigma_0^2$ 时，似然函数 $L(\mu, \sigma_0^2) = (2\pi)^{-\frac{n}{2}}\sigma_0^{-n} e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2}$，

有 $\ln L(\mu, \sigma_0^2) = -\dfrac{n}{2}\ln(2\pi) - \dfrac{n}{2}\ln\sigma_0^2 - \dfrac{1}{2\sigma_0^2}\sum\limits_{i=1}^{n}(x_i-\mu)^2$，

令 $\dfrac{d\ln L}{d\mu}=-\dfrac{1}{2\sigma_0^2}\sum_{i=1}^{n}2(x_i-\mu)\cdot(-1)=\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=\dfrac{1}{\sigma^2}\left(\sum_{i=1}^{n}x_i-n\mu\right)=0$，得 $\mu=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$，

则 $\sup\limits_{\mu}L(\mu,\sigma_0^2)=(2\pi)^{-\frac{n}{2}}\sigma_0^{-n}\,\mathrm{e}^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\bar{x})^2}=(2\pi)^{-\frac{n}{2}}\sigma_0^{-n}\,\mathrm{e}^{-\frac{(n-1)s^2}{2\sigma_0^2}}$，

故似然比检验统计量为

$$\Lambda(X_1,\cdots,X_n)=\frac{\sup\limits_{\mu,\sigma^2}L(\mu,\sigma^2)}{\sup\limits_{\mu}L(\mu,\sigma_0^2)}=\left[\frac{(n-1)S^2}{n\sigma_0^2}\right]^{-\frac{n}{2}}\mathrm{e}^{-\frac{n}{2}+\frac{(n-1)S^2}{2\sigma_0^2}},$$

这与统计量 $\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$ 相对应.

3． 设 $X_1,\cdots,X_n$ 为来自指数分布 $Exp(\lambda_1)$的样本，$Y_1,\cdots,Y_m$ 为来自指数分布 $Exp(\lambda_2)$的样本，且两组样本独立，其中λ_1,λ_2是未知的正参数.

（1）求假设 H_0：$\lambda_1=\lambda_2$ vs H_1：$\lambda_1\neq\lambda_2$ 的似然比检验；

（2）证明上述检验法的拒绝域仅依赖于比值 $\sum_{i=1}^{n}X_i\Big/\sum_{i=1}^{m}Y_i$；

（3）求统计量 $\sum_{i=1}^{n}X_i\Big/\sum_{i=1}^{m}Y_i$ 在原假设成立下的分布.

解：（1）因样本联合密度函数为

$$p(x_1,\cdots,x_n,y_1,\cdots,y_m;\lambda_1,\lambda_2)=\prod_{i=1}^{n}\lambda_1\mathrm{e}^{-\lambda_1x_i}\prod_{i=1}^{m}\lambda_2\mathrm{e}^{-\lambda_2y_i}=\lambda_1^n\lambda_2^m\,\mathrm{e}^{-\lambda_1\sum_{i=1}^{n}x_i-\lambda_2\sum_{i=1}^{m}y_i},$$

则似然函数 $L(\lambda_1,\lambda_2)=\lambda_1^n\lambda_2^m\,\mathrm{e}^{-\lambda_1\sum_{i=1}^{n}x_i-\lambda_2\sum_{i=1}^{m}y_i}$， $\ln L(\lambda_1,\lambda_2)=n\ln\lambda_1+m\ln\lambda_2-\lambda_1\sum_{i=1}^{n}x_i-\lambda_2\sum_{i=1}^{m}y_j$，

令 $\begin{cases}\dfrac{\partial\ln L}{\partial\lambda_1}=\dfrac{n}{\lambda_1}-\sum\limits_{i=1}^{n}x_i=0;\\[2mm] \dfrac{\partial\ln L}{\partial\lambda_2}=\dfrac{n}{\lambda_2}-\sum\limits_{i=1}^{m}y_i=0.\end{cases}$

得 $\lambda_1=\dfrac{n}{\sum\limits_{i=1}^{n}x_i}$， $\lambda_2=\dfrac{m}{\sum\limits_{i=1}^{m}y_i}$，

则 $\sup\limits_{\lambda_1,\lambda_2}L(\lambda_1,\lambda_2)=\dfrac{n^nm^m}{\left(\sum\limits_{i=1}^{n}x_i\right)^n\left(\sum\limits_{i=1}^{m}y_i\right)^m}\mathrm{e}^{-n-m}$，

当$\lambda_1=\lambda_2$时，似然函数 $L(\lambda_1)=\lambda_1^{n+m}\,\mathrm{e}^{-\lambda_1\left(\sum_{i=1}^{n}x_i+\sum_{i=1}^{m}y_i\right)}$， $\ln L(\lambda_1)=(n+m)\ln\lambda_1-\lambda_1\left(\sum_{i=1}^{n}x_i+\sum_{i=1}^{m}y_j\right)$，

令 $\dfrac{d\ln L(\lambda_1)}{d\lambda_1}=\dfrac{n+m}{\lambda_1}-\left(\sum_{i=1}^n x_i+\sum_{i=1}^m y_j\right)=0$，得 $\lambda_1=\dfrac{n+m}{\sum\limits_{i=1}^n x_i+\sum\limits_{i=1}^m y_i}$，

则 $\sup\limits_{\lambda_1=\lambda_2} L(\lambda_1,\lambda_2)=\dfrac{(n+m)^{n+m}}{\left(\sum\limits_{i=1}^n x_i+\sum\limits_{i=1}^m y_i\right)^{n+m}}\mathrm{e}^{-n-m}$，

故似然比检验统计量为

$$\Lambda(X_1,\cdots,X_n,Y_1,\cdots,Y_m)=\frac{\sup\limits_{\lambda_1,\lambda_2}L(\lambda_1,\lambda_2)}{\sup\limits_{\lambda_1=\lambda_2}L(\lambda_1,\lambda_2)}=\frac{n^n m^n\left(\sum\limits_{i=1}^n X_i+\sum\limits_{i=1}^m Y_i\right)^{n+m}}{(n+m)^{n+m}\left(\sum\limits_{i=1}^n X_i\right)^n\left(\sum\limits_{i=1}^m Y_i\right)^m}$$

$$=\frac{n^n m^n}{(n+m)^{n+m}}\left(\frac{\sum\limits_{i=1}^n X_i+\sum\limits_{i=1}^m Y_i}{\sum\limits_{i=1}^n X_i}\right)^n\cdot\left(\frac{\sum\limits_{i=1}^n X_i+\sum\limits_{i=1}^m Y_i}{\sum\limits_{i=1}^m Y_i}\right)^m=\frac{n^n m^n}{(n+m)^{n+m}}\left(1+\frac{\sum\limits_{i=1}^m Y_i}{\sum\limits_{i=1}^n X_i}\right)^n\cdot\left(1+\frac{\sum\limits_{i=1}^n X_i}{\sum\limits_{i=1}^m Y_i}\right)^m;$$

（2）因似然比检验统计量 $\Lambda(X_1,\cdots,X_n,Y_1,\cdots,Y_m)=\dfrac{n^n m^n}{(n+m)^{n+m}}\left(1+\sum\limits_{i=1}^m Y_i\Big/\sum\limits_{i=1}^n X_i\right)^n\cdot\left(1+\sum\limits_{i=1}^n X_i\Big/\sum\limits_{i=1}^m Y_i\right)^m$，

故拒绝域仅依赖于比值 $\sum\limits_{i=1}^n X_i\Big/\sum\limits_{i=1}^m Y_i$；

（3）因 $X_i\sim Exp(\lambda_1)$，有 $2\lambda_1 X_i\sim Exp\left(\dfrac{1}{2}\right)=Ga\left(1,\dfrac{1}{2}\right)=\chi^2(2)$，且 $X_1,\cdots,X_n$ 相互独立，

则 $2\lambda_1\sum\limits_{i=1}^n X_i\sim\chi^2(2n)$，同理 $2\lambda_2\sum\limits_{i=1}^m Y_i\sim\chi^2(2m)$，

因两组样本独立，

故 $F=\dfrac{2\lambda_1\sum\limits_{i=1}^n X_i\Big/(2n)}{2\lambda_2\sum\limits_{i=1}^m Y_i\Big/(2m)}=\dfrac{m}{n}\cdot\dfrac{\sum\limits_{i=1}^n X_i}{\sum\limits_{i=1}^m Y_i}\sim F(2n,2m)$.

4. 设 $X_1,\cdots,X_n$ 为来自正态分布 $N(\mu,\sigma^2)$的 i.i.d.样本，其中μ,σ^2 未知. 证明关于假设 $H_0:\mu\le\mu_0$ vs $H_1:\mu>\mu_0$ 的单侧 t 检验是似然比检验（显著性水平$\alpha<1/2$）.

证：因样本联合密度函数为

$$p(x_1,\cdots,x_n;\mu,\sigma^2)=\prod_{i=1}^n\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=(2\pi)^{-\frac{n}{2}}\sigma^{-n}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^n(x_i-\mu)^2},$$

则似然函数 $L(\mu,\sigma^2)=(2\pi)^{-\frac{n}{2}}\sigma^{-n}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^n(x_i-\mu)^2}$，

有 $\ln L(\mu,\sigma^2)=-\dfrac{n}{2}\ln(2\pi)-\dfrac{n}{2}\ln\sigma^2-\dfrac{1}{2\sigma^2}\sum\limits_{i=1}^n(x_i-\mu)^2$，

令$\begin{cases}\dfrac{\partial \ln L}{\partial \mu} = -\dfrac{1}{2\sigma^2}\sum\limits_{i=1}^{n} 2(x_i-\mu)\cdot(-1) = \dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu) = \dfrac{1}{\sigma^2}\left(\sum\limits_{i=1}^{n} x_i - n\mu\right) = 0; \\ \dfrac{\partial \ln L}{\partial \sigma^2} = -\dfrac{n}{2}\cdot\dfrac{1}{\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\sum\limits_{i=1}^{n}(x_i-\mu)^2 = 0. \end{cases}$

得 $\mu = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i = \bar{x}$， $\sigma^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\mu)^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$，

则 $\sup\limits_{\mu,\sigma^2} L(\mu,\sigma^2) = (2\pi)^{-\frac{n}{2}}\left[\dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\right]^{-\frac{n}{2}} e^{-\frac{n}{2}} = (2\pi)^{-\frac{n}{2}}\left[\dfrac{(n-1)s^2}{n}\right]^{-\frac{n}{2}} e^{-\frac{n}{2}}$，

当$\mu \le \mu_0$时，若 $\bar{x} \le \mu_0$，有 $\sup\limits_{\mu\le\mu_0,\sigma^2} L(\mu,\sigma^2) = \sup\limits_{\mu,\sigma^2} L(\mu,\sigma^2)$，

则似然比检验统计量 $\Lambda(X_1,\cdots,X_n) = \dfrac{\sup\limits_{\mu,\sigma^2} L(\mu,\sigma^2)}{\sup\limits_{\mu\le\mu_0,\sigma^2} L(\mu,\sigma^2)} = 1$，

若 $\bar{x} > \mu_0$，似然函数上确界应在$\mu = \mu_0$时取得，

即似然函数 $L(\sigma^2) = (2\pi)^{-\frac{n}{2}}\sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu_0)^2}$，有 $\ln L(\sigma^2) = -\dfrac{n}{2}\ln(2\pi)^{-\frac{n}{2}} - \dfrac{n}{2}\ln\sigma^2 - \dfrac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu_0)^2$，

令 $\dfrac{d\ln L}{d(\sigma^2)} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu_0)^2 = 0$，

得 $\sigma^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\mu_0)^2 = \dfrac{1}{n}\left[\sum\limits_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu_0)^2\right] = \dfrac{(n-1)s^2}{n} + (\bar{x}-\mu_0)^2$，

则 $\sup\limits_{\mu\le\mu_0,\sigma^2} L(\mu,\sigma^2) = (2\pi)^{-\frac{n}{2}}\left[\dfrac{(n-1)s^2}{n} + (\bar{x}-\mu_0)^2\right]^{-\frac{n}{2}} e^{-\frac{n}{2}}$，

故似然比检验统计量为

$$\Lambda(X_1,\cdots,X_n) = \frac{\sup\limits_{\mu,\sigma^2} L(\mu,\sigma^2)}{\sup\limits_{\mu\le\mu_0,\sigma^2} L(\mu,\sigma^2)} = \left[\frac{(n-1)S^2}{n}\right]^{-\frac{n}{2}}\left[\frac{(n-1)S^2}{n} + (\bar{X}-\mu_0)^2\right]^{\frac{n}{2}}$$

$$= \left[1 + \frac{1}{n-1}\left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right)^2\right]^{\frac{n}{2}},$$

这与关于假设 H_0：$\mu \le \mu_0$ vs H_1：$\mu > \mu_0$ 的单侧 t 检验的统计量 $T = \dfrac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$ 相对应.

5. 按孟德尔遗传规律，让开淡红花的豌豆随机交配，子代可区分为红花、淡红花和白花三类，且其比例是 1:2:1，为了验证这个理论，观察一次实验，得到红花、淡红花和白花的豌豆株数分别为 26, 66, 28，这些数据与孟德尔定律是否一致（$\alpha = 0.05$）？

解：假设 H_0： $p_1=\frac{1}{4}, p_2=\frac{1}{2}, p_3=\frac{1}{4}$，

选取统计量 $\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}\stackrel{\sim}{\cdot}\chi^2(r-1)$，

显著性水平 $\alpha=0.05$，$r=3$， $\chi^2_{1-\alpha}(r-1)=\chi^2_{0.95}(2)=5.9915$，右侧拒绝域 $W=\{\chi^2\ge 5.9915\}$，

因 $n=120$， p_i, n_i 及计算结果如下表：

花色	红花	淡红花	白花	合计
n_i	26	66	28	120
p_i	1/4	1/2	1/4	1
n_i-np_i	−4	6	−2	0
$(n_i-np_i)^2/(np_i)$	0.5333	0.6	0.1333	1.2666

有 $\chi^2=1.2666\notin W$，并且检验的 p 值 $p=P\{\chi^2\ge 1.2666\}=0.5308>\alpha=0.05$，
故接受 H_0，拒绝 H_1，即可以认为这些数据与孟德尔定律一致.

6. 掷一颗骰子 60 次，结果如下：

点数	1	2	3	4	5	6
次数	7	8	12	11	9	13

试在显著性水平为 0.05 下检验这颗骰子是否均匀.

解：假设 H_0： $p_1=p_2=\cdots=p_6=\frac{1}{6}$，

选取统计量 $\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}\stackrel{\sim}{\cdot}\chi^2(r-1)$，

显著性水平 $\alpha=0.05$，$r=6$， $\chi^2_{1-\alpha}(r-1)=\chi^2_{0.95}(5)=11.0705$，右侧拒绝域 $W=\{\chi^2\ge 11.0705\}$，

因 $n=60$，p_i, n_i 及计算结果如下表：

点数	1	2	3	4	5	6	合计
n_i	7	8	12	11	9	13	60
p_i	1/6	1/6	1/6	1/6	1/6	1/6	1
n_i-np_i	−3	−2	2	1	−1	3	0
$(n_i-np_i)^2/(np_i)$	0.9	0.4	0.4	0.1	0.1	0.9	2.8

有 $\chi^2=2.8\notin W$，并且检验的 p 值 $p=P\{\chi^2\ge 2.8\}=0.7308>\alpha=0.05$，
故接受 H_0，拒绝 H_1，即可以认为这颗骰子是均匀的.

7. 检查了一本书的 100 页，记录各页中的印刷错误的个数，其结果如下：

错误个数	0	1	2	3	4	5	≥6
页数	35	40	19	3	2	1	0

问能否认为一页的印刷错误个数服从泊松分布（取 $\alpha=0.05$）？

解：为了使得在每一类中的个数不小于 5，将取值为 3、4、5、≥6 的情形合并为一类，即

错误个数	0	1	2	≥3
页数	35	40	19	6

假设 H_0： $p_i=\frac{\lambda^i}{i!}e^{-\lambda}$， $i=0,1,2$ 且 $p_3=\sum_{i=3}^{+\infty}\frac{\lambda^i}{i!}e^{-\lambda}$，

需估计一个参数λ，$k=1$，选取统计量 $\chi^2=\sum_{i=0}^{r-1}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i} \dot{\sim} \chi^2(r-k-1)$，

显著性水平$\alpha=0.05$，$r=4$，$\chi^2_{1-\alpha}(r-k-1)=\chi^2_{0.95}(2)=5.9915$，右侧拒绝域 $W=\{\chi^2\ge 5.9915\}$，

因 $n=100$，$\hat{\lambda}=\bar{x}=\frac{100}{100}=1$，$\hat{p}_i, n_i$ 及计算结果如下表：

错误个数	0	1	2	3	合计
n_i	35	40	19	6	100
$\hat{p}_i$	0.3679	0.3679	0.1839	0.0803	1
$n_i-n\hat{p}_i$	−1.7879	3.2120	0.6060	−2.0301	0
$(n_i-n\hat{p}_i)^2/(n\hat{p}_i)$	0.0869	0.2804	0.0200	0.5132	0.9005

有$\chi^2=0.9005\notin W$，并且检验的 p 值 $p=P\{\chi^2\ge 0.9005\}=0.6375>\alpha=0.05$，
故接受 H_0，拒绝 H_1，即可以认为一页的印刷错误个数服从泊松分布.

8. 某建筑工地每天发生事故数现场记录如下：

一天发生的事故数	0	1	2	3	4	5	≥6	合计
天数	102	59	30	8	0	1	0	200

试在显著性水平$\alpha=0.05$ 下检验这批数据是否服从泊松分布.

解：为了使得在每一类中的个数不小于 5，将取值为 3、4、5、≥6 的情形合并为一类，即

假设 H_0：$p_i=\frac{\lambda^i}{i!}e^{-\lambda}$，$i=0,1,2$ 且 $p_3=\sum_{i=3}^{+\infty}\frac{\lambda^i}{i!}e^{-\lambda}$，

需估计一个参数λ，$k=1$，选取统计量 $\chi^2=\sum_{i=0}^{r-1}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i} \dot{\sim} \chi^2(r-k-1)$，

显著性水平$\alpha=0.05$，$r=4$，$\chi^2_{1-\alpha}(r-k-1)=\chi^2_{0.95}(2)=5.9915$，右侧拒绝域 $W=\{\chi^2\ge 5.9915\}$，

因 $n=200$，$\hat{\lambda}=\bar{x}=\frac{148}{200}=0.74$，$\hat{p}_i, n_i$ 及计算结果如下表：

一天发生的事故数	0	1	2	≥3	合计
n_i	102	59	30	9	200
$\hat{p}_i$	0.4771	0.3531	0.1306	0.0392	1
$n_i-n\hat{p}_i$	6.5772	−11.6129	3.8732	1.1624	0
$(n_i-n\hat{p}_i)^2/(n\hat{p}_i)$	0.4533	1.9098	0.5742	0.1724	3.1097

有$\chi^2=3.1097\notin W$，并且检验的 p 值 $p=P\{\chi^2\ge 3.1097\}=0.2112>\alpha=0.05$，
故接受 H_0，拒绝 H_1，即可以认为这批数据服从泊松分布.

9. 在一批灯泡中抽取 300 只作寿命试验，其结果如下：

寿命（h）	< 100	[100, 200)	[200, 300)	≥ 300
灯泡数	121	78	43	58

在显著性水平为 0.05 下能否认为灯泡寿命服从指数分布 $Exp(0.005)$？

解：假设 H_0：$p_i=e^{-100(i-1)\lambda}-e^{-100i\lambda}$，$i=1,2,3$ 且 $p_4=e^{-300\lambda}$，

选取统计量 $\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i} \dot{\sim} \chi^2(r-1)$，

显著性水平$\alpha = 0.05$，$r = 4$， $\chi^2_{1-\alpha}(r-1) = \chi^2_{0.95}(3) = 7.8147$，右侧拒绝域 $W = \{\chi^2 \ge 7.8147\}$，

因 $n = 300$，$\lambda = 0.005$，p_i, n_i 及计算结果如下表：

寿命（h）	< 100	$[100, 200)$	$[200, 300)$	≥ 300	合计
n_i	121	78	43	58	300
p_i	0.3935	0.2387	0.1447	0.2231	1
$n_i - np_i$	2.9592	6.4046	−0.4248	−8.9390	0
$(n_i - np_i)^2/(np_i)$	0.0742	0.5729	0.0042	1.1937	1.8450

有$\chi^2 = 1.8450 \notin W$，并且检验的 p 值 $p = P\{\chi^2 \ge 1.8450\} = 0.6052 > \alpha = 0.05$，
故接受 H_0，拒绝 H_1，即可以认为灯泡寿命服从指数分布 $Exp(0.005)$.

10．下表是上海 1875 年到 1955 年的 81 年间，根据其中 63 年观察到的一年中（5 月到 9 月）下暴雨次数的整理资料

i	0	1	2	3	4	5	6	7	8	≥ 9
n_i	4	8	14	19	10	4	2	1	1	0

试检验一年中暴雨次数是否服从泊松分布（$\alpha = 0.05$）？

解：将取值为 0、1 的情形合并为一类，并将 5、6、7、8、≥9 的情形合并为一类，即

i	≤ 1	2	3	4	≥ 5
n_i	12	14	19	10	8

假设 H_0：$p_1 = \sum_{i=0}^{1} \frac{\lambda^i}{i!} e^{-\lambda}$， $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$， $i = 2, 3, 4$ 且 $p_5 = \sum_{i=5}^{+\infty} \frac{\lambda^i}{i!} e^{-\lambda}$，

需估计一个参数λ，$k = 1$，选取统计量 $\chi^2 = \sum_{i=0}^{r-1} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} \dot{\sim} \chi^2(r-k-1)$，

显著性水平$\alpha = 0.05$，$r = 5$， $\chi^2_{1-\alpha}(r-k-1) = \chi^2_{0.95}(3) = 7.8147$，右侧拒绝域 $W = \{\chi^2 \ge 7.8147\}$，

因 $n = 100$，$\hat{\lambda} = \bar{x} = \frac{180}{63} = 2.8571$， $\hat{p}_i, n_i$ 及计算结果如下表：

暴雨次数	1	2	3	4	5	合计
n_i	12	14	19	10	8	63
$\hat{p}_i$	0.2215	0.2344	0.2233	0.1595	0.1613	1
$n_i - n\hat{p}_i$	−1.9566	−0.7684	4.9349	−0.0465	−2.1634	0
$(n_i - n\hat{p}_i)^2/(n\hat{p}_i)$	0.2743	0.0400	1.7314	0.0002	0.4605	2.5064

有$\chi^2 = 2.5064 \notin W$，并且检验的 p 值 $p = P\{\chi^2 \ge 2.5064\} = 0.5440 > \alpha = 0.05$，
故接受 H_0，拒绝 H_1，即可以认为上海一年中暴雨次数服从泊松分布.

11．某种配偶的后代按体格的属性分为三类，各类的数目分别是 10, 53, 46．按照某种遗传模型其频率之比应为 $p^2 : 2p(1-p) : (1-p)^2$，问数据与模型是否相符（$\alpha = 0.05$）？

解：假设 H_0：$p_1 = p^2$，$p_2 = 2p(1-p)$，$p_3 = (1-p)^2$，

需估计一个参数 p，$k = 1$，选取统计量 $\chi^2 = \sum_{i=1}^{r} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} \dot{\sim} \chi^2(r-k-1)$，

显著性水平$\alpha = 0.05$，$r = 3$， $\chi^2_{1-\alpha}(r-k-1) = \chi^2_{0.95}(1) = 3.8415$，右侧拒绝域 $W = \{\chi^2 \ge 3.8415\}$，

设后代的各类数目分别为 n_1, n_2, n_3 次，有 $n_1 + n_2 + n_3 = n$，

则似然函数 $L(p)=(p^2)^{n_1}[2p(1-p)]^{n_2}[(1-p)^2]^{n_3}=2^{n_2}p^{2n_1+n_2}(1-p)^{n_2+2n_3}$，

有 $\ln L(p)=n_2\ln 2+(2n_1+n_2)\ln p+(n_2+2n_3)\ln(1-p)$，

令 $\dfrac{d\ln L(p)}{dp}=(2n_1+n_2)\cdot\dfrac{1}{p}+(n_2+2n_3)\cdot\dfrac{-1}{1-p}=0$，

得 p 的 MLE $\hat{p}=\dfrac{2n_1+n_2}{2(n_1+n_2+n_3)}=\dfrac{2n_1+n_2}{2n}$，

因 $n=109$，$\hat{p}=\dfrac{2n_1+n_2}{2n}=\dfrac{73}{218}=0.3349$，$\hat{p}_i, n_i$ 及计算结果如下表：

后代类别	1	2	3	合计
n_i	10	53	46	109
$\hat{p}_i$	0.1121	0.4455	0.4424	1
$n_i-n\hat{p}_i$	−2.2225	4.4450	−2.2225	0
$(n_i-n\hat{p}_i)^2/(n\hat{p}_i)$	0.4041	0.4069	0.1024	0.9135

有 $\chi^2=0.9135\notin W$，并且检验的 p 值 $p=P\{\chi^2\ge 0.9135\}=0.6608>\alpha=0.05$，故接受 H_0，拒绝 H_1，即可以认为数据与模型相符.

12．按有无特性 A 与 B 将 n 个样品分成四类，组成 2×2 列联表：

	B	$\overline{B}$	合计
A	a	b	$a+b$
$\overline{A}$	c	d	$c+d$
合计	$a+c$	$b+d$	n

其中 $n=a+b+c+d$，试证明此时列联表独立性检验的 χ^2 统计量可以表示成

$$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

证：假设 H_0：$p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$，$i=1, 2$；$j=1, 2$，

选取统计量 $\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{2}\dfrac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}\sim\chi^2(1)$，

因 $\hat{p}_{1\cdot}=\dfrac{a+b}{n}$，$\hat{p}_{2\cdot}=\dfrac{c+d}{n}$，$\hat{p}_{\cdot 1}=\dfrac{a+c}{n}$，$\hat{p}_{\cdot 2}=\dfrac{b+d}{n}$，且 $n=a+b+c+d$，

则 $\hat{p}_{11}=\hat{p}_{1\cdot}\cdot\hat{p}_{\cdot 1}=\dfrac{(a+b)(a+c)}{n^2}$，$\hat{p}_{12}=\dfrac{(a+b)(b+d)}{n^2}$，$\hat{p}_{21}=\dfrac{(c+d)(a+c)}{n^2}$，$\hat{p}_{22}=\dfrac{(c+d)(b+d)}{n^2}$，

$$\text{故 }\chi^2=\frac{\left[a-\frac{(a+b)(a+c)}{n}\right]^2}{\frac{(a+b)(a+c)}{n}}+\frac{\left[b-\frac{(a+b)(b+d)}{n}\right]^2}{\frac{(a+b)(b+d)}{n}}+\frac{\left[c-\frac{(c+d)(a+c)}{n}\right]^2}{\frac{(c+d)(a+c)}{n}}+\frac{\left[d-\frac{(c+d)(b+d)}{n}\right]^2}{\frac{(c+d)(b+d)}{n}}$$

$$=\frac{[na-(a+b)(a+c)]^2}{n(a+b)(a+c)}+\frac{[nb-(a+b)(b+d)]^2}{n(a+b)(b+d)}+\frac{[nc-(c+d)(a+c)]^2}{n(c+d)(a+c)}+\frac{[nd-(c+d)(b+d)]^2}{n(c+d)(b+d)}$$

$$=\frac{(ad-bc)^2}{n(a+b)(a+c)}+\frac{(bc-ad)^2}{n(a+b)(b+d)}+\frac{(bc-ad)^2}{n(c+d)(a+c)}+\frac{(ad-bc)^2}{n(c+d)(b+d)}$$

$$=\frac{(ad-bc)^2[(c+d)(b+d)+(c+d)(a+c)+(a+b)(b+d)+(a+b)(a+c)]}{n(a+b)(c+d)(a+c)(b+d)}$$

$$=\frac{[(a+b)+(c+d)][(b+d)+(a+c)](ad-bc)^2}{n(a+b)(c+d)(a+c)(b+d)}$$

$$=\frac{n^2(ad-bc)^2}{n(a+b)(c+d)(a+c)(b+d)}$$

$$=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

13．在研究某种新措施对猪白痢的防治效果问题时，获得了如下数据：

	存活数	死亡数	合计	死亡率
对照	114	36	150	24%
新措施	132	18	150	12%
合计	246	54	300	36%

试问新措施对防治该种疾病是否有显著疗效（$\alpha = 0.05$）？

解：假设 H_0：$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$，$i = 1, 2$；$j = 1, 2$，

选取统计量 $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}} \dot{\sim} \chi^2(1)$，

显著性水平 $\alpha = 0.05$，$\chi^2_{1-\alpha}(1) = \chi^2_{0.95}(1) = 3.8415$，右侧拒绝域 $W = \{\chi^2 \ge 3.8415\}$，

因 $n = 300$，$\hat{p}_{1\cdot} = \frac{150}{300} = 0.5$，$\hat{p}_{2\cdot} = \frac{150}{300} = 0.5$，$\hat{p}_{\cdot 1} = \frac{246}{300} = 0.82$，$\hat{p}_{\cdot 2} = \frac{54}{300} = 0.18$，

且 $\hat{p}_{ij} = \hat{p}_{i\cdot} \cdot \hat{p}_{\cdot j}$，$i, j = 1, 2$，$n_{ij}$ 及计算结果如下表：

措施	对照		新措施		合计
效果	存活	死亡	存活	死亡	
n_{ij}	114	36	132	18	300
$\hat{p}_{ij}$	0.41	0.09	0.41	0.09	1
$n_{ij}-n\hat{p}_{ij}$	−9	9	9	−9	0
$(n_i-n\hat{p}_i)^2/(n\hat{p}_i)$	0.6585	3	0.6585	3	7.3170

有 $\chi^2 = 7.3170 \in W$，并且检验的 p 值 $p = P\{\chi^2 \ge 7.3170\} = 0.0068 < \alpha = 0.05$，

故拒绝 H_0，接受 H_1，即可以认为新措施对防治该种疾病有显著疗效.

14．某单位调查了 520 名中年以上的脑力劳动者，其中 136 人有高血压史，另外 384 人则无．在有高血压史的 136 人中，经诊断为冠心病及可疑者的有 48 人，在无高血压史的 384 人中，经诊断为冠心病及可疑者的有 36 人．从这个资料，对高血压与冠心病有无关系作检验，取 $\alpha = 0.01$．

解：假设 H_0：$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$，$i = 1, 2$；$j = 1, 2$，

选取统计量 $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}} \dot{\sim} \chi^2(1)$，

显著性水平 $\alpha = 0.01$，$\chi^2_{1-\alpha}(1) = \chi^2_{0.99}(1) = 6.6349$，右侧拒绝域 $W = \{\chi^2 \ge 6.6349\}$，

因 $n=520$，$\hat{p}_{1\cdot}=\dfrac{136}{520}=0.2615$，$\hat{p}_{2\cdot}=\dfrac{384}{520}=0.7385$，$\hat{p}_{\cdot 1}=\dfrac{48+36}{520}=0.1615$，$\hat{p}_{\cdot 2}=\dfrac{88+348}{300}=0.8385$，

且 $\hat{p}_{ij}=\hat{p}_{i\cdot}\cdot\hat{p}_{\cdot j}$， $i,j=1,2$，n_{ij} 及计算结果如下表：

血压	高		低		合计
冠心病	有	无	有	无	
n_{ij}	48	88	36	348	520
$\hat{p}_{ij}$	0.0422	0.2193	0.1193	0.6192	1
$n_{ij}-n\hat{p}_{ij}$	26.0308	–26.0308	–26.0308	26.0308	
$(n_i-n\hat{p}_i)^2/(n\hat{p}_i)$	30.8432	5.9423	10.9236	2.1046	49.8136

有 $\chi^2=49.8136\in W$，并且检验的 p 值 $p=P\{\chi^2\ge 49.8136\}=1.6906\times 10^{-12}<\alpha=0.05$，

故拒绝 H_0，接受 H_1，即可以认为高血压与冠心病有关系．

15．在一项是否应提高小学生的计算机课程的比例的调查结果如下：

年龄	同意	不同意	不知道
55 岁以上	32	28	14
36 ~ 55 岁	44	21	17
15 ~ 35 岁	47	12	13

问年龄因素是否影响了对问题的回答（$\alpha=0.05$）？

解：假设 H_0：$p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$，$i=1,2,3$；$j=1,2,3$，

选取统计量 $\chi^2=\displaystyle\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}\ \dot{\sim}\ \chi^2(4)$，

显著性水平 $\alpha=0.05$， $\chi^2_{1-\alpha}(4)=\chi^2_{0.95}(4)=9.4877$，右侧拒绝域 $W=\{\chi^2\ge 9.4877\}$，

因 $n=228$， $\hat{p}_{1\cdot}=\dfrac{32+28+14}{228}=0.3246$， $\hat{p}_{2\cdot}=\dfrac{44+21+17}{228}=0.3596$， $\hat{p}_{3\cdot}=\dfrac{47+12+13}{228}=0.3158$，

$$\hat{p}_{\cdot 1}=\frac{32+44+47}{228}=0.5395,\quad \hat{p}_{\cdot 2}=\frac{28+21+12}{228}=0.2675,\quad \hat{p}_{\cdot 3}=\frac{14+17+13}{228}=0.1930,$$

且 $\hat{p}_{ij}=\hat{p}_{i\cdot}\cdot\hat{p}_{\cdot j}$， $i,j=1,2$，n_{ij} 及计算结果如下表：

年龄	55 岁以上			36 ~ 55 岁			15 ~ 35 岁			合计
回答	同意	不同意	不知道	同意	不同意	不知道	同意	不同意	不知道	
n_{ij}	32	28	14	44	21	17	47	12	13	228
$\hat{p}_{ij}$	0.1751	0.0868	0.0626	0.1940	0.0962	0.0694	0.1704	0.0845	0.0609	1
$n_{ij}-n\hat{p}_{ij}$	–7.9211	8.2018	–0.2807	–0.2368	–0.9386	1.1754	8.1579	–7.2632	–0.8947	0
$(n_i-n\hat{p}_i)^2/(n\hat{p}_i)$	1.5717	3.3977	0.0055	0.0013	0.0402	0.0873	1.7134	2.7386	0.0576	9.6132

有 $\chi^2=9.6132\in W$，并且检验的 p 值 $p=P\{\chi^2\ge 9.6132\}=0.0475<\alpha=0.05$，

故拒绝 H_0，接受 H_1，即可以认为年龄因素影响了对问题的回答．