RSS Higher Certificate in Statistics, 2010

Module 5: Further Probability and Inference

Solutions

Question 1

(i) $P(X=0,Y=0)=P(Y=0\mid X=0)P(X=0)=1\times 0.4=0.4$

$P(X=1,Y=0)=P(Y=0\mid X=1)P(X=1)=0.6\times 0.3=0.18$

$P(X=1,Y=1)=P(Y=1\mid X=1)P(X=1)=0.4\times 0.3=0.12$

$P(X=2,Y=0)=P(Y=0\mid X=2)P(X=2)=0.6^2\times 0.2=0.072$

$P(X=2,Y=1)=P(Y=1\mid X=2)P(X=2)=2\times 0.6\times 0.4\times 0.2=0.096$

$P(X=2,Y=2)=P(Y=2\mid X=2)P(X=2)=0.4^2\times 0.2=0.032$

$P(X=3,Y=0)=P(Y=0\mid X=3)P(X=3)=0.6^3\times 0.1=0.0216$

$P(X=3,Y=1)=P(Y=1\mid X=3)P(X=3)=3\times 0.6^2\times 0.4\times 0.1=0.00432$

$P(X=3,Y=2)=P(Y=2\mid X=3)P(X=3)=3\times 0.6\times 0.4^2\times 0.1=0.0288$

$P(X=3,Y=3)=P(Y=3\mid X=3)P(X=3)=0.4^3\times 0.1=0.0064$

2 元分割表の他のすべてのマス内は 0 となる．もちろん，明らかに上のすべてを計算する必要はない．多くは引き算で求められるからだ．

表は以下のようになる．

		Values of X				
		0	1	2	3	Total
Values of Y	0	0.4	0.18	0.072	0.0216	0.6736
	1	0	0.12	0.096	0.0432	0.2592
	2	0	0	0.032	0.0288	0.0608
	3	0	0	0	0.0064	0.0064
	Total	0.4	0.3	0.2	0.1	

(ii) Y の周辺分布は(i)の表の Total 列に等しい．

$P(Y=0)=0.6736$，$P(Y=1)=0.2592$，$P(Y=2)=0.0608$，$P(Y=3)=0.0064$

$\therefore E(Y)=0\times 0.6736+1\times 0.2592+2\times 0.0608+3\times 0.0064=0.4$

(iii) $Cov(X,Y)=E(XY)-E(X)E(Y)$．表より

$E(X)=0\times 0.4+1\times 0.3+2\times 0.2+3\times 0.1=1.0,$

$E(XY)=0\times 0\times 0.4+......+3\times 3\times 0.0064=0.8$

$\therefore Cov(X,Y)=0.8-1.0\times 0.4=0.4$

(iv)　ここで U を $X-Y$ とし，U の確率分布を求めると，

$P(U=0)=P(X=Y)=0.4+0.12+0.032+0.064=0.5584$

$P(U=1)=P(X-Y=1)=0.18+0.096+0.0288=0.3048$

$P(U=2)=P(X-Y=2)=0.072+0.0432=0.1152$

$P(U=3)=P(X-Y=3)=0.0216$

[$k=0,1,2,3$ でないすべての k について $P(U=k)=0$]

Question 2

$$f(x,y)=\frac{1}{2\pi}\exp\left(-\frac{1}{4}(x-1)^2-(y-\frac{1}{4}(1+x))^2\right),\quad -\infty<x<\infty,-\infty<y<\infty$$

(i) $$f\left(x\right)=\int_{y=-\infty}^{\infty}f\left(x,y\right)dy=\frac{1}{2\pi}\exp\left(-\frac{1}{4}(x-1)^2\right)\int_{y=-\infty}^{\infty}\exp\left(-(y-\frac{1}{4}(1+x))^2\right)dy$$

この積分を正規確率密度関数の積分として表す

$$=\frac{1}{2\pi}\exp\left(-\frac{1}{4}(x-1)^2\right)\sqrt{\pi}\int_{y=-\infty}^{\infty}\underline{\underline{\frac{1}{\sqrt{2\pi\times\frac{1}{2}}}\exp\left(-\frac{\left(y-\frac{1}{4}(1+x)\right)^2}{2\times\frac{1}{2}}\right)}}dy$$

下線部が確率密度関数 $N\left(\frac{1+x}{4},\frac{1}{2}\right)$ であることに注意

$$=\frac{1}{\sqrt{2\pi\times 2}}\exp\left(-\frac{\left(x-1\right)^2}{2\times 2}\right)\quad\left(-\infty<x<\infty\right)$$

これは N(1, 2)の確率密度関数だから，X は N(1, 2)に従う.

(ii) X の積率母関数は

$$m_X\left(t\right)=\int_{-\infty}^{\infty}e^{tx}f\left(x\right)dx=\frac{1}{\sqrt{4\pi}}\int_{-\infty}^{\infty}e^{tx}e^{-\frac{1}{4}(x-1)^2}dx$$

$$=\frac{1}{\sqrt{4\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{4}\left(x^2-2x+1-4tx\right)}dx$$

$$=\frac{1}{\sqrt{4\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{4}\left(x-(1+2t)\right)^2}e^{\frac{1}{4}\left((1+2t)^2-1\right)}dx$$

$$=e^{t+t^2}\int_{-\infty}^{\infty}\underline{\underline{\frac{1}{\sqrt{2\pi\times 2}}e^{-\frac{1}{4}\left(x-(1+2t)\right)^2}}}dx\quad=e^{t+t^2}$$

下線部が確率密度関数 N(1+2t, 2)であることに注意

(iii) 積率母関数を微分して，

$$\frac{dm_X(t)}{dt}=\left(1+2t\right)e^{t+t^2}$$

$$\frac{d^2m_X(t)}{dt^2}=\left(1+2t\right)^2e^{t+t^2}+2e^{t+t^2}$$

$$\frac{d^3m_X(t)}{dt^3}=\left(1+2t\right)^3e^{t+t^2}+4\left(1+2t\right)e^{t+t^2}+2\left(1+2t\right)e^{t+t^2}$$

$$\therefore E\left(X^3\right)=\left.\frac{d^3m_X(t)}{dt^3}\right|_{t=0}=1+4+2=7$$

[注：積率母関数のべき級数展開を用いて容易に結果を得ることもできる．]

Question 3

(i)　尤度関数は $L(\beta)=\prod_{i=1}^{n}\left(\beta(1-x_i)^{\beta-1}\right)=\beta^n\prod_{i=1}^{n}(1-x_i)^{\beta-1}$

$\therefore \log L(\beta)=n\log\beta-(\beta-1)\sum\log(1-x_i)$

$\frac{d\log L}{d\beta}=\frac{n}{\beta}-\sum\log(1-x_i)$　これを 0 とおいて β について解くと，$\hat{\beta}=-\frac{n}{\sum\log(1-x_i)}$

これが最大となることは $\frac{d^2\log L}{d\beta^2}=-\frac{n}{\beta^2}<0$ で確認されるので，この $\hat{\beta}$ が β の最尤推定量である.

(ii)　$E\left(-\frac{d^2\log L}{d\beta^2}\right)=\frac{n}{\beta^2}$ なので，通常の近似式を用いると $Var\left(\hat{\beta}\right)\approx\frac{1}{n/\beta^2}=\frac{\beta^2}{n}$

よって大きい n では，$\hat{\beta}\sim N\left(\beta,\frac{\beta^2}{n}\right)$ と近似できる.

近似的な β の 95%信頼区間は $\hat{\beta}\pm1.96\frac{\hat{\beta}}{\sqrt{n}}$ で与えられる.

[注：pivotal quantity 法を用いて他の値を求めることもできる.]

(iii)　$P(X_i<0.5)=\int_0^{0.5}\beta(1-x)^{\beta-1}dx=\left[-(1-x)^{\beta}\right]_0^{0.5}=-0.5^{\beta}+1=1-0.5^{\beta}$

(iv)　$Y\sim B\left(n,1-0.5^{\beta}\right)$ だから，

$P(Y=y)=\binom{n}{y}\left(1-0.5^{\beta}\right)^{y}\left(0.5^{\beta}\right)^{n-y}$

尤度関数は $\binom{n}{y}\left(1-0.5^{\beta}\right)^{y}\left(0.5^{\beta}\right)^{n-y}$ である.

Question 4

(a) パラメータ θ の推定量 $\hat{\theta}$ のバイアスは，$E\left(\hat{\theta}-\theta\right)$ と定義される．

不偏推定量とは，全ての θ に対して $E\left(\hat{\theta}-\theta\right)=0$ の成り立つ推定量のことである．

推定量のバイアスが 0 でなく，異なる標本から計算される各推定値が独立である場合，標本サイズにかかわらず，これらの推定量の平均値は真の値 θ に依存しない．ゆえに，他の条件が等しければ，不偏推定量の中から決定するのが好まれる．だが，偏りのある推定量が最良の不偏推定量よりも小さな分散を持っている状況があるため，（少なくともバイアスが小さいときには）その偏りのある推定量を採用するべきだ．実際，いくつかの状況では不偏推定量が存在しない．

不偏推定量の中で最適なのは，最小の分散を持つものである．1 つの不偏推定量 $\hat{\theta}_1$ に対する，別の不偏推定量 $\hat{\theta}_2$ と比較した相対的有効性は

$$\frac{\mathrm{Var}\left(\hat{\theta}_2\right)}{\mathrm{Var}\left(\hat{\theta}_1\right)}$$

である（ときどき，%表示のためこれに 100 をかけたものを扱う）．1(100%)より大きい値だと $\hat{\theta}_1$ の方がよく，1 より小さい値だと $\hat{\theta}_2$ の方がよい．

正則条件下では，不偏推定量の分散は，L を尤度関数としたとき

$$-\frac{1}{E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)} \quad (=V\text{とする})$$

未満にはなりえないことがわかる．不偏推定量 $\hat{\theta}$ の有効性は $\frac{V}{\mathrm{Var}\left(\hat{\theta}\right)}$ である（もしくは，%表示のためこれに 100 をかけたもの）．この値が 1(100%)のとき $\hat{\theta}$ が最もよい不偏推定量だということを表している．

(b) (i) 各 Y_i を考えて

$$E(Y)=(-2)\times\frac{1}{2}(1-p)^2+0\times p(1-p)+1\times p+2\times\frac{1}{2}(1-p)^2=p$$

よってモーメント推定量 $\tilde{p}$ は $\bar{Y}=\tilde{p}$ を満たす．つまり $\tilde{p}$ は単に $\bar{Y}$ でよい．

p は 0 から 1 の間の値でなければならないが，$\tilde{p}$ の値はこの範囲外になりうる，という性質は明らかに不適切である．

(ii)

$$E\left(Y_i^2\right)=4\times\frac{1}{2}(1-p)^2+0\times p(1-p)+1\times p+4\times\frac{1}{2}(1-p)^2$$

$$=4(1-p)^2+p=4p^2-7p+4$$

$$\therefore \mathrm{Var}\left(Y_i^2\right)=E\left(Y_i^2\right)-\left\{E\left(Y_i\right)\right\}^2=4p^2-7p+4-p^2=3p^2-7p+4$$

$$\therefore \mathrm{Var}\left(\tilde{p}\right)=\mathrm{Var}\left(\overline{Y}\right)=\frac{3p^2-7p+4}{n}$$

(iii) (ii)より， $E\left(\frac{1}{n}\sum Y_i^2\right)=4p^2-7p+4$， $E\left(\frac{1}{4n}\sum Y_i^2\right)=p^2-\frac{7}{4}p+1$

$$\therefore E\left(\frac{1}{4n}\sum Y_i^2+\frac{7}{4n}\sum Y_i-1\right)=p^2$$

よって $\frac{1}{4n}\sum Y_i^2+\frac{7}{4n}\sum Y_i-1$ が p^2 の不偏推定量である．