Homework#1 Statistical Inference II

Name: Jia-Han Shih

Problem 3.1 [p.134]

(a) In Example 3.1, show that $\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{T(n-T)}{n}$.

Solution:

According to Example 3.1, we have

$$X_1, X_2, \cdots, X_n \overset{iid}{\sim} \text{Ber}(p), \quad 0 < p < 1.$$

Then $T = \sum_{i=1}^{n} X_i$ is a complete sufficient statistics. Since

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = E\left(\frac{T}{n}\right) = p$$

Hence T/n is the UMVUE of p. Now,

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2) &= \sum_{i=1}^{n} X_i^2 - 2\overline{X}\sum_{i=1}^{n} X_i + n\overline{X}^2 \\ &= \sum_{i=1}^{n} X_i^2 - 2n\overline{X}^2 + n\overline{X}^2 \\ &= \sum_{i=1}^{n} X_i^2 - n\overline{X}^2.\end{aligned}$$

Since $X_1, X_2, \cdots, X_n$ follow Bernoulli distribution, we have

$$\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i.$$

Thus, the equality become

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2 = \sum_{i=1}^{n} X_i - n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 &= T - \frac{T^2}{n} \\ &= \frac{T(n-T)}{n}.\end{aligned}$$

Then we finished the proof.

(b) The variance of $\frac{T(n-T)}{n(n-1)}$ in Example 3.1 is $\frac{pq}{n}\left\{(q-p)^2+\frac{2pq}{n-1}\right\}$, where $q=1-p$.

Solution:

By direct calculation, we have

$$\begin{aligned} E\left(\frac{T(n-T)}{n(n-1)}\right) &= \sum_{t=0}^{n}\frac{n!}{t!(n-t)!}\cdot\frac{t(n-t)}{n(n-1)}p^t(1-p)^{n-t} \\ &= p(1-p)\sum_{t=1}^{n-1}\frac{(n-2)!}{(t-1)!(n-t-1)!}p^{t-1}(1-p)^{n-t-1} \\ &= p(1-p). \end{aligned}$$

Since $\frac{T(n-T)}{n(n-1)}$ is a function of complete sufficient statistics. Therefore, $\frac{T(n-T)}{n(n-1)}$ is the UMVUE of $p(1-p)$. The variance is

$$\begin{aligned} &\operatorname{var}\left\{\frac{T(n-T)}{n(n-1)}\right\} \\ &= E\left\{\frac{T^2(n-T)^2}{n^2(n-1)^2}\right\} - E\left\{\frac{T(n-T)}{n(n-1)}\right\}^2 \\ &= \frac{1}{n^2(n-1)^2}\{E(T^4)-2nE(T^3)+n^2E(T^2)-n^2(n-1)^2p^2(1-p)^2\}. \end{aligned}$$

To calculate the variance, we need to first compute the moments of Binomial distributions. The moment generating function of Binomial distribution is

$$M(t)=(1-p+pe^t)^n.$$

Then we can compute the moments by moment generating function. The first moment is

$$E(T)=\frac{d}{dt}M(t)\bigg|_{t=0}=npe^t(1-p+pe^t)^{n-1}\Big|_{t=0}=np.$$

The second moment is

$$
\begin{aligned}
E(T^2) &= \frac{d^2}{dt^2} M(t)\Big|_{t=0} \\
&= \{ npe^t (1-p+pe^t)^{n-1} + n(n-1)p^2 e^{2t} (1-p+pe^t)^{n-2} \}_{t=0} \\
&= np + n(n-1)p^2 = n^2 p^2 - np^2 + np.
\end{aligned}
$$

The third moment is

$$
\begin{aligned}
E(T^3) &= \frac{d^3}{dt^3} M(t)\Big|_{t=0} \\
&= \{ npe^t (1-p+pe^t)^{n-1} + n(n-1)p^2 e^{2t} (1-p+pe^t)^{n-2} \\
&\quad + 2n(n-1)p^2 e^{2t} (1-p+pe^t)^{n-2} + n(n-1)(n-2)p^3 e^{3t} (1-p+pe^t)^{n-2} \}_{t=0} \\
&= np + n(n-1)p^2 + 2n(n-1)p^2 + n(n-1)(n-2)p^3 \\
&= n^3 p^3 - 3n^2 p^3 + np^3 + 3n^2 p^2 - 3np^2 + np.
\end{aligned}
$$

The fourth moment is

$$
\begin{aligned}
E(T^4) &= \frac{d^4}{dt^4} M(t)\Big|_{t=0} \\
&= \{ npe^t (1-p+pe^t)^{n-1} + n(n-1)p^2 e^{2t} (1-p+pe^t)^{n-2} \\
&\quad + 2n(n-1)p^2 e^{2t} (1-p+pe^t)^{n-2} + n(n-1)(n-2)p^3 e^{3t} (1-p+pe^t)^{n-2} \\
&\quad + 4n(n-1)p^2 e^{2t} (1-p+pe^t)^{n-2} + 2n(n-1)(n-2)p^3 e^{3t} (1-p+pe^t)^{n-2} \\
&\quad + 3n(n-1)(n-2)p^3 e^{3t} (1-p+pe^t)^{n-2} \\
&\quad + n(n-1)(n-2)(n-3)p^4 e^{4t} (1-p+pe^t)^{n-3} \}_{t=0} \\
&= np + n(n-1)p^2 + 2n(n-1)p^2 + n(n-1)(n-2)p^3 \\
&\quad + 4n(n-1)p^2 + 2n(n-1)(n-2)p^3 + 3n(n-1)(n-2)p^3 \\
&\quad + n(n-1)(n-2)(n-3)p^4 \\
&= n^4 p^4 - 6n^3 p^4 + 11n^2 p^4 - 6np^4 + 6n^3 p^3 - 18n^2 p^3 + 12np^3 + 7n^2 p^2 - 7np^2 + np.
\end{aligned}
$$

Therefore, the variance is

$$\operatorname{var}\left\{\frac{T(n-T)}{n(n-1)}\right\}$$

$$=\frac{1}{n^2(n-1)^2}\{E(T^4)-2nE(T^3)+n^2E(T^2)-n^2(n-1)^2p^2(1-p)^2\}$$

$$=\frac{1}{n^2(n-1)^2}$$

$$\times\{n^4p^4-6n^3p^4+11n^2p^4-6np^4+6n^3p^3-18n^2p^3+12np^3+7n^2p^2-7np^2+np$$

$$-2n^4p^3+6n^3p^3-4n^2p^3-6n^3p^2+6n^2p^2-2n^2p+n^4p^2-n^3p^2+n^3p$$

$$-n^4p^4+2n^3p^4-n^2p^4+2n^4p^3-4n^3p^3+2n^2p^3-n^4p^2+2n^3p^2-n^2p^2\}$$

$$=\frac{1}{n^2(n-1)^2}\{-4n^3p^4+10n^2p^4-6np^4+8n^3p^3-20n^2p^3+12np^3$$

$$-5n^3p^2+12n^2p^2-7np^2+n^3p^2-2n^2p+np\}$$

$$=\frac{p}{n(n-1)^2}\{-4n^2p^3+10np^3-6p^3+8n^2p^2-20np^2+12p^2$$

$$-5n^2p+12np-7p+n^2-2n+1\}$$

$$=\frac{p}{n(n-1)^2}\{-2(n-1)(2n-3)p^3+4(n-1)(2n-3)p^2$$

$$-(n-1)(5n-7)p+(n-1)^2\}$$

$$=\frac{p}{n(n-1)}\{6p^3-4np^3+8np^2-12p^2+7p-5np+n-1\}$$

$$=\frac{p}{n(n-1)}\{n(-4p^3+8p^2-5p+1)+4p^3-8p^2+5p-1+2p^3-4p^2+2p\}$$

$$=\frac{p}{n(n-1)}\{n(1-2p)^2(1-p)-(1-2p)^2(1-p)+2p(1-p)^2\}$$

$$=\frac{p}{n(n-1)}\{(n-1)(1-2p)^2(1-p)+2p(1-p)^2\}$$

$$=\frac{p(1-p)}{n}\left\{(1-2p)^2+\frac{2p(1-p)}{n-1}\right\}$$

$$=\frac{pq}{n}\left\{(q-p)^2+\frac{2pq}{n-1}\right\}.$$

Then we finished the proof.

Problem 1.6

In Example 1.5, find the Bayes estimator δ_Λ of $p(1-p)$ when p has the prior $B(a,b)$.

Solution:

Suppose

$$X \mid p \sim \text{Bin}(n, p) \quad \text{and} \quad p \sim \text{Beta}(a,b),$$

where $a,b>0$. Then we have

$$\begin{aligned}\pi(p \mid x) &\propto p^{x}(1-p)^{n-x}p^{a-1}(1-p)^{b-1}\\ &\propto p^{x+a-1}(1-p)^{n-x+b-1}.\end{aligned}$$

Therefore, the posterior distribution is

$$p \mid x \sim \text{Beta}(x+a, n-x+b).$$

The Bayes estimator δ_Λ of $p(1-p)$ is

$$\begin{aligned}\delta_\Lambda &= E\{p(1-p) \mid x\}\\ &= \int_0^1 p(1-p)\frac{\Gamma(a+b+n)}{\Gamma(x+a)\Gamma(n-x+b)}p^{x+a-1}(1-p)^{n-x+b-1}dp\\ &= \int_0^1 \frac{\Gamma(a+b+n)}{\Gamma(x+a)\Gamma(n-x+b)}p^{x+a+1-1}(1-p)^{n-x+b+1-1}dp\\ &= \frac{(x+a)(n-x+b)}{(a+b+n)(a+b+n+1)}\\ &\quad \times\int_0^1 \frac{(a+b+n)(a+b+n+1)\Gamma(a+b+n)}{(x+a)\Gamma(x+a)(n-x+b)\Gamma(n-x+b)}p^{x+a+1-1}(1-p)^{n-x+b+1-1}dp\\ &= \frac{(x+a)(n-x+b)}{(a+b+n)(a+b+n+1)}\int_0^1 \frac{\Gamma(a+b+n+2)}{\Gamma(x+a+1)\Gamma(n-x+b+1)}p^{x+a+1-1}(1-p)^{n-x+b+1-1}dp\\ &= \frac{(x+a)(n-x+b)}{(a+b+n)(a+b+n+1)}.\end{aligned}$$

Then we obtain the Bayes estimator

$$\delta_\Lambda = \frac{(x+a)(n-x+b)}{(a+b+n)(a+b+n+1)}.$$

Example 2.5

Suppose

$$X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(0, \sigma^2).$$

Then we have

$$\begin{aligned} f(x|\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-x_i^2}{2\sigma^2} \right\} \\ &\propto \left(\frac{1}{\sigma^2} \right)^{n/2} e^{-\frac{\Sigma x_i^2}{2\sigma^2}}. \end{aligned}$$

Letting $\tau = \frac{1}{2\sigma^2}$ and $r = \frac{n}{2}$, the joint density can be rewrite as

$$f(x|\sigma^2) \propto \tau^r e^{-r\Sigma x_i^2}.$$

This representation is similar to the Gamma distribution. Hence we can assume a conjugate prior

$$\tau \sim \text{Gamma}(g, 1/\alpha).$$

By the properties of the Gamma distribution, we have

$$E(\tau) = \frac{g}{\alpha}, \quad E(\tau^2) = \frac{g}{\alpha^2} + \frac{g^2}{\alpha^2} = \frac{g(g+1)}{\alpha^2},$$

$$E\left(\frac{1}{\tau}\right) = \frac{\alpha}{g-1} \text{ and } E\left(\frac{1}{\tau^2}\right) = \frac{\alpha^2}{(g-1)(g-2)}.$$

Then we can derive the posterior distribution, let $Y = \sum_{i=1}^{n} X_i^2$. Thus,

$$\begin{aligned} \pi(\tau|x) &\propto \tau^r e^{-\tau y} \tau^{g-1} e^{-\alpha y} \\ &\propto \tau^{r+g-1} e^{-(y+\alpha)\tau}. \end{aligned}$$

Therefore, the posterior distribution is

$$\pi(\tau|x) \sim \text{Gamma}(r+g, 1/(y+\alpha)).$$

The Bayes estimator of σ^2 is

$$\delta_\Lambda = E(\sigma^2 \mid x) = E(1/2\tau \mid x)$$
$$= \frac{1}{2} E(\tau \mid x) = \frac{1}{2} \cdot \frac{y+\alpha}{r+g-1} = \frac{y+\alpha}{n+2g-2}.$$

As we known, the UMVUE of σ^2 is

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{Y}{n}$$

and the prior mean is

$$E(1/2\tau) = \frac{1}{2} E(1/\tau) = \frac{1}{2} \cdot \frac{\alpha}{g-1} = \frac{\alpha}{2g-2}.$$

Therefore, the Bayes estimator can be written as a linear combination of the UMVUE and the prior mean

$$\delta_\Lambda = \frac{y+\alpha}{n+2g-2} = \frac{n}{n+2g-2} \cdot \frac{y}{n} + \frac{2g-2}{n+2g-2} \cdot \frac{\alpha}{2g-2}$$
$$= m \cdot \frac{y}{n} + (1-m)\frac{\alpha}{2g-2},$$

where $m = \dfrac{n}{n+2g-2}$.