Homework 1: Review of materials

Solution Set

Problem 1: Prove that the Binomial distribution arises as a sum of n iid Bernoulli trials each with success probability p.

We prove the above using moment-generating functions.

Suppose, we have i.i.d. Bernouli random variables, $X_1, ..., X_n \sim Ber(p)$. These RVs share the MGF

$$M_{X_i}(t) = (1-p) + pe^t.$$

Define their sum $S_n = \sum_1^n X_i$

Recall that the sum of n i.i.d. random variables has MGF equal to the nˆth power of the MGF of the individual variable, therefore the MGF for the sum of n Bernoulli trials is

$$M_{S_n}(t) = ((1-p) + pe^t)^n = \sum_{r=0}^{n} \binom{n}{r} (pe^t)^r (1-p)^{n-r}.$$

Here, the last equality follows due to Binomial theorem. On the other hand, the moment-generating function of a Binomial distribution, $Y \sim Bin(n, p)$ is

$$M_Y(t) = \sum_{r=0}^{n} e^{rt} \binom{n}{r} p^r (1-p)^{n-r}. = \sum_{r=0}^{n} \binom{n}{r} (pe^t)^r (1-p)^{n-r}.$$

Thus, the MGFs are same, Hence the proof.

Problem 2: Let $l(\theta)$ denote a twice continuously differentiable log likelihood corresponding to an iid sample under density f_θ where n is the sample size. The score function is defined as

$$u(\theta) = \frac{\partial l(\theta)}{\partial \theta},$$

and the Fisher information matrix is defined as

$$I(\theta) = -\mathrm{E}\left(\frac{\partial^2 l(\theta)}{\partial \theta^2}\right),$$

where the expectation is over the assumed distribution for the data when the parameter value is θ. Prove that

$$E(u(\theta)) = 0 \qquad \textbf{and} \qquad Var(u(\theta)) = I(\theta).$$

The score function is,

$$u(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \log[\prod_{i=1}^{n} f_\theta(x_i)] = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f_\theta(x_i) = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial \theta} f_\theta(x_i)}{f_\theta(x_i)}$$

So we have,

$$
\begin{aligned}
E(u(\theta)) &= E(\frac{\partial l(\theta)}{\partial \theta}) \\
&= \sum \int_{\chi} \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} f_\theta(x) dx \\
&= \sum \int_{\chi} \frac{\partial}{\partial \theta} f_\theta(x) dx \\
&= \sum \frac{\partial}{\partial \theta} \int_{\chi} f_\theta(x) dx \\
&= \sum \frac{\partial}{\partial \theta} 1 \\
&= 0
\end{aligned}
$$

ii)

By the results we just proved in **i)**,

$$
\begin{aligned}
0 &= E(u(\theta)) \\
0 &= \frac{\partial}{\partial \theta} E(u(\theta)) \\
&= \frac{\partial}{\partial \theta} E(\sum \frac{\partial}{\partial \theta} \log f_\theta(x)) \\
&= \frac{\partial}{\partial \theta} \int_{\chi} \sum \frac{\partial \log f_\theta(x)}{\partial \theta} f_\theta(x) dx \\
&= \sum \int_{\chi} \frac{\partial}{\partial \theta} [\frac{\partial \log f_\theta(x)}{\partial \theta} f_\theta(x)] dx \\
&= \sum \int_{\chi} [\frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} f_\theta(x) + \frac{\partial \log f_\theta(x)}{\partial \theta} \frac{\partial f_\theta(x)}{\partial \theta}] dx \\
&= \sum \int_{\chi} \frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} f_\theta(x) dx + \sum \int_{\chi} \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} f_\theta(x) dx
\end{aligned}
$$

Rearrange the above equality to get,

$$
\begin{aligned}
-\sum \int_{\chi} \frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} f_\theta(x) dx &= \sum \int_{\chi} \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} f_\theta(x) dx \\
-E(\frac{\partial^2 l(\theta)}{\partial \theta^2}) &= E[(\frac{\partial l(\theta)}{\partial \theta})^2] \\
I(\theta) &= E[u(\theta)]^2
\end{aligned}
$$

Finally, note that $E[u(\theta)] = 0$, hence we arrive at,

$$
I(\theta) = E[u(\theta)]^2 = E[u(\theta) - E(u(\theta))]^2 = Var(u(\theta))
$$

Problem 3: Let $Y \sim$ binomial(n, π) and let $T_n = \hat{\pi} = Y/n$. Use the CLT and the Delta Method to construct an asymptotic confidence interval for logit(π). Note that this recipe does not work when the estimated success probability is on the boundary of its support, ie $\hat{\pi} = 0$ or $\hat{\pi} = 1$. Why?

From Problem 1, we can write Y as the sum of n i.i.d. varibles X_i with Bernoulli distribution with probability π. i.e. $Y = \sum_{i=1}^{n} X_i$

By CLT, we have $\sqrt{n}(T_n - \pi) \to \mathcal{N}(0, \pi(1-\pi))$in distribution, as $g(x) = logit(x)$ fulfills $g'(\pi) = \frac{1}{\pi(1-\pi)} \neq 0 (0 < \pi < 1)$, we can obtain the following results with Delta Methods.

$$\sqrt{n}(g(T_n) - g(\pi)) \to \mathcal{N}(0, \pi(1-\pi)(g'(\pi))^2)$$
$$\sqrt{n}(g(T_n) - g(\pi)) \to \mathcal{N}(0, \frac{1}{\pi(1-\pi)})$$

If $0 < \pi < 1$, we can also get $T_n \to \pi$ in probability by WLLN which indicates $\sqrt{\frac{1}{\pi(1-\pi)}} \to \sqrt{\frac{1}{\hat{\pi}(1-\hat{\pi})}}$ in probability. From this, we can build asymptotic $1-\alpha$ confidence interval for $logit(\pi)$ as

$$\left[\log\big(\frac{\hat{\pi}}{1-\hat{\pi}}\big) - z_{\alpha/2}\sqrt{\frac{1}{n\hat{\pi}(1-\hat{\pi})}}, \log\big(\frac{\hat{\pi}}{1-\hat{\pi}}\big) + z_{\alpha/2}\sqrt{\frac{1}{n\hat{\pi}(1-\hat{\pi})}}\right]$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$ perventile of the standard normal distribution.

To explain why this cannot work for $\hat{\pi} = 0$ ro 1, we can see that the derivative of $g(\theta)$ does not exist at 0 and 1. This indicates that we cannot apply Delta Method to build the interval and the recipe does not work due to this.