Definition 6.20 (efficiency, efficient, asymptotically efficient, TBp.302)

- The **efficiency** of an unbiased estimator $\hat{\theta}$ of θ is

cf. relative efficiency (LNp.59) — a function of θ

$$\text{eff}_\theta(\hat{\theta}) = \frac{1/nI_{X_1}(\theta)}{\text{Var}_\theta(\hat{\theta})} = \frac{1}{nI_{X_1}(\theta)\text{Var}_\theta(\hat{\theta})}.$$

$\boxed{\text{eff}_\theta(\hat{\theta})=1}$ ★ $\boxed{0 \le \text{eff}_\theta(\hat{\theta}) \le 1}$

- An unbiased estimator whose variance achieves the lower bound is called **efficient**.

asymptotic efficiency $\equiv \lim_{n\to\infty} \text{eff}_\theta(\hat{\theta}_n)$

1. consistent ($\hat{\theta}_n \xrightarrow{P} \theta$) $\Rightarrow$ asymptotically unbiased ($E(\hat{\theta}_n)\to\theta$)
2. Usually (check LNp.40) bias$^2 \sim o(1/n)$, var $\sim O(1/n)$ $\Rightarrow$ MSE $\approx$ var

why no "unbiased"

- If the asymptotic variance of an estimator equals the lower bound, cf. the estimator is said to be **asymptotically efficient**. $\boxed{\lim_{n\to\infty} \text{eff}_\theta(\hat{\theta}_n)=1}$

Notes (TBp.302)

1. MLE is asymptotically efficient. ($\because$ asymptotic variance of MLE $= \frac{1}{nI_X(\theta)}$)
2. For a finite sample size, MLE may not be efficient. Thm6.6 (LNp.39)
3. MLEs are not the only asymptotically efficient estimators. shrinkage
4. There may exist biased and super efficient estimators. For example, if $X_1, \dots, X_n$ is an iid sample $\sim$ Normal$(\mu, 1)$. Let $\tilde{\mu} = \overline{X}\, I(|\overline{X}| > n^{-1/4})$. Then

MLE: $\hat{\mu}=\overline{X}$

Note. By CLT, $\sqrt{n}(\overline{X}-\mu) \xrightarrow{d} N(0,1)$

$n^{-1/2}$ converges to 0 much faster than $n^{-1/4}$

$$\sqrt{n}(\tilde{\mu}-\mu) \xrightarrow{D} N(0, \sigma^2(\mu)) \quad \Bigg| = \begin{cases} \overline{X}, & \text{if } |\overline{X}|>n^{-1/4} \\ 0, & \text{if } |\overline{X}|\le n^{-1/4} \end{cases}$$

where $\sigma^2(\mu) = 1$ if $\mu \neq 0$ and $\sigma^2(0) = 0$. $\tilde{\mu}$ much more efficient than $\hat{\mu}$ (MLE) at $\mu=0$

Example 6.30 (Poisson distribution, TBp.302)

For the Poisson distribution, $I_{X_1}(\lambda) = 1/\lambda$. For any unbiased estimator T of λ, based on a sample of i.i.d. Poisson variables $X_1, X_2, \dots, X_n$, statistical modeling

LNp.42, Ex.6-19

$$\text{Var}_\lambda(T) \ge \lambda/n. \leftarrow \text{C.-R. bound}$$

MLE of λ is $\overline{X} = S/n$ and $S=\sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda) \Rightarrow E(\overline{X}) = \frac{n\lambda}{n} = \lambda \leftarrow$ unbiased

LNp.19, Ex.6-10

$$\text{Var}_\lambda(\overline{X}) = \lambda/n.$$

$\text{eff}_\lambda(\overline{X})=1 \Rightarrow \overline{X}$ is efficient $\forall\lambda$.

Hence $\overline{X}$ attains the Cramer-Rao lower bound. (**Note.** There may exist a biased estimator of λ that has a smaller MSE than $\overline{X}$.)

Theorem 6.16 (generalization of Cramer-Rao inequality)

Suppose $X_1, \dots, X_n$ are i.i.d. with pdf/pmf $f(x|\theta)$, and $T = T(X_1, \dots, X_n)$ is an unbiased estimator of $\tau(\theta)$. Then, under smooth assumptions on $f(x|\theta)$,

$E_\theta(T)=\tau(\theta)$

$$\text{Var}_\theta(T) \ge \frac{\tau'(\theta)^2}{nI_{X_1}(\theta)} = \frac{1}{n\cdot\left(\frac{I_{X_1}(\theta)}{\tau'(\theta)^2}\right)}$$

cf. Thm. 6.7 & 6.8, LNp.41~42

NB(k,p) in LN.CH1~8, p.63 cf.

Example 6.31 (generalized negative binomial, TBp.302-305)

- The generalized negative binomial distribution has the pmf: $\frac{x_i}{k} \approx \frac{m}{k}$

Note. m & k are parameters

$$f(x|m,k) = \left(\frac{k}{m+k}\right)^k \frac{\Gamma(k+x)}{x!\Gamma(k)} \left(\frac{m}{m+k}\right)^x, \quad x = 0, 1, \dots.$$

$m=(1-p)k/p$; $\frac{k}{m+k} = p$; $\frac{m}{m+k} = 1-p$; $m = k/p - k$

Note that its mean is $\mu = m$ and its variance is $\sigma^2 = m + m^2/k \leftarrow \frac{k(1-p)}{p^2}$

unknown: m : # of black balls, k : # of white balls $\Longrightarrow$ X = # of black balls that will be drawn before the kth white ball is drawn (draw with replacement)

failure, 0; success, 1; prob of getting white: $p=\frac{k}{m+k}$

- Generalized negative binomial model arises in several cases.
  1. If k is an integer, then the number of failures up to the kth success in a sequence of Bernoulli trails with probability of success $p = k/(m + k)$ follows a generalized negative binomial distribution.

     textbook CH4, Problem 92

  2. Hierarchical model: If $\Lambda \sim \text{Gamma}(k, k/m)$, and $X|\Lambda = \lambda \sim \text{Poisson}(\lambda)$, then $X \sim$ generalized negative binomial. ← marginal distribution
  3. Biological model (Solomon, 1983): If N, the number of clusters/colonies, follows a Poisson(λ) distribution and X_i, $i = 1, \ldots, N$, the numbers of individuals in each colonies, follows a logarithmic series distribution with success probability p

     $X = X_1 + X_2 + \cdots + X_N$ (random)

     $$P(X_i = t) = \frac{-1}{\log(p)} \frac{(1-p)^t}{t}, \quad t = 1, 2, \cdots.$$

     x: an individual; a cluster

     Assume that the X_i's are independent. Then the distribution of the total number of individuals is generalized negative binomial with $k = -\lambda/\log(p)$ and $p = m/(m + k)$.

     ① Find cond'al mgf of $X|N$. ② Use law of total expectation to find the mgf of X: $Ee^{tX}=EE[e^{tX}|N]$

  4. birth, death rate = constant; immigration rate = constant
     Population size ∼ generalized negative binomial

     birth & death process

- Estimation of m and k (Note. both m & $k>0$)

  statistical modeling: $X_1, \cdots, X_n \overset{i.i.d.}{\sim} GNB(m,k)$

  method 1: method of moments estimators

  $$\hat{m} = \overline{X}, \quad \hat{k} = \frac{\overline{X}^2}{\hat{\sigma}^2 - \overline{X}}$$

  reasonable?

  $\left.\begin{array}{l}\mu = m\\ \sigma^2 = m + \frac{m^2}{k}\end{array}\right\} \Rightarrow \left\{\begin{array}{l}m=\mu\\ k=\frac{\mu^2}{\sigma^2-\mu}\end{array}\right.$

  method 2:

  statistic → n_0 = number of zeros (i.e., # of $X_i=0$) out of a sample of size n → $\frac{n_0}{n} \approx P(X_i=0) \equiv p_0$

  parameter → $p_0 = \left(\frac{k}{m+k}\right)^k$ = probability of zero count

  $n_0 \sim B(n, p_0)$

  If m is estimated by $\overline{X}$, then k can be estimated by solving the equation

  $$p_0 \approx \frac{n_0}{n} = \left(\frac{\hat{k}}{\hat{k} + \overline{X}}\right)^{\hat{k}} = \left(1 + \frac{\overline{X}}{\hat{k}}\right)^{-\hat{k}}.$$

  Compared to method 3 (MLE)

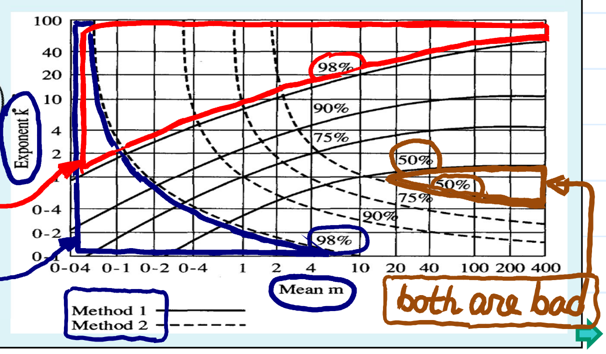
  no close form ⇒ use asymptotic variance of MLE ← numerical method

  method 3: MLE is difficult to compute. MLE of m is the sample mean $\overline{X}$, but MLE of k is the solution of a nonlinear equation.

  MLE is the best

- Asymptotic relative efficiencies of estimators (**Figure 8.11 in textbook**)
  — Method 2 is quite efficient when p_0 close to 1 or 0. (reasonable?)
  — Method 1 becomes more efficient as k increases.

  method 1 is good; method 2 is good



- **Insect counts data, Bliss and Fisher (1953)**
  - 25 leaves from each of 6 apple trees in an sprayed orchard
  - number of adult female red mites on each of the leaves

X_i's

Data: $X_1, \cdots, X_{25}$ ← 1st tree; $X_{26}, \cdots, X_{50}$ ← 2nd tree; ⋮ ; $X_{126}, \cdots, X_{150}$ ← 6th tree; i.i.d. X

number per leaf	observed count	Poisson fit	Negative Binomial fit
0	70	47.7	69.5
1	38	54.6	37.6
2	17	31.3	20.1
3	10	12.0	10.7
4	9	3.4	5.7
5	3	0.75	3.0
6	2	0.15	1.6
7	1	0.03	0.85
8^+	0	0.00	0.95

$\hat{P}(X_i=0)=\frac{70}{150}$ histogram; information lost; $X \sim P(\lambda)$; $P(X=s|\hat{\lambda})$; $\overline{X}=\hat{\lambda}$: suff. + complete; $X \sim NB(m,k)$; $P(X=s|\hat{m},\hat{k})$; s

- Recursive algorithm for pmf of generalized negative binomial:

$$f(0|m,k) = \left(1+\frac{m}{k}\right)^{-k}, \quad f(n|m,k) = \frac{k+n-1}{n}\left(\frac{m}{k+m}\right) f(n-1|m,k)$$

Recall. 2 in LNp.66

- Poisson fit is not good. Heterogeneous: the rates of insect infection might be different on different trees and at different locations on the same tree. (For the Poisson fit, polling the last 3 cells: $\chi^2 = 82.4$, df=5, extremely small p-value; For the Negative Binomial fit, pooling last 2 cells: $\chi^2 = 2.48$, df=5, p-value=0.22)

must use information not contained in the suff. stat. $\hat{\lambda}$

❖ **Reading**: textbook, 8.7; **Further reading**: Hogg et al., 6.2

• method of finding estimator III --- UMVUE (or MVUE)

Question 6.7 ("best" unbiased estimator)

Among *all* unbiased estimators, how to find the "best" estimator, i.e., the one that has the smallest MSE?

(**Recall.** [1]. $MSE_\theta(\hat{\theta}) = Var_\theta(\hat{\theta})$ for any unbiased estimator $\hat{\theta}$;

[2]. Cramer-Rao bound) ← LNp.62 (& efficient)

An unbiased estimator whose variance reaches the C.-R. bound is a UMVUE.

Definition 6.21 (UMVUE)

An estimator T^* of $\tau(\theta)$ is called a **(uniformly) minimum variance unbiased estimator (UMVUE or MVUE)** of $\tau(\theta)$ if

1. T^* is unbiased for $\tau(\theta)$, and
2. for any unbiased estimator T of $\tau(\theta)$, $\text{Var}_\theta(T^*) \le \text{Var}_\theta(T)$ for all $\theta \in \Omega$.

LNp.65

Example 6.32 (cont. Ex. 6.30, UMVUE of Poisson mean, TBp.302)

- The Cramer-Rao bound for $\tau(\theta) = \theta$ is θ/n.
- On the other hand, $\text{E}(\overline{X}) = \theta$ and $\text{Var}(\overline{X}) = \theta/n$.
- Hence, $\overline{X}$ is a UMVUE of θ. MLE and unbiased

See Ex.6.35 (LNp.73)

Question

A UMVUE may not achieve the Cramer-Rao lower bound. Are there other *systematic procedures* that can be used to derive/identify a UMVUE among a large number of unbiased estimators?

Theorem 6.17 (Rao-Blackwell theorem, TBp.310)

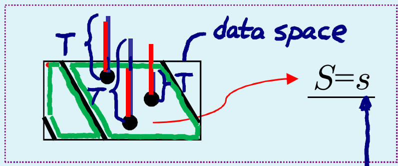
Suppose $X_1, \ldots, X_n$ have a joint pdf or pmf $f(x_1, \ldots, x_n; \theta)$, and

$$S = (S_1, \ldots, S_k)$$

is sufficient for θ. If T is any estimator of $\tau(\theta)$ with $\mathrm{E}(T^2) < \infty$, let

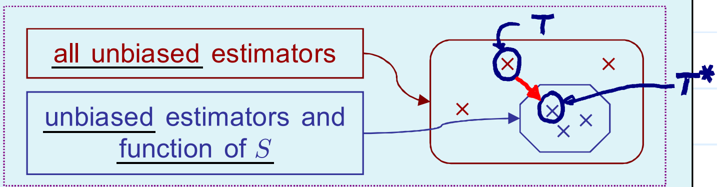
$$T^* \stackrel{\downarrow}{=} \mathrm{E}_\theta(\, T \mid S \,).$$

Q: What if ① $S=(X_1,\cdots,X_n)$? ② S is minimal sufficient?



the value of S might not be directly used to estimate θ (Ex6.28 in LNp.57)

① a function of S ② the values of T^* come from the values of T (cf. value of S)



Then, i.e., T & T^* have same bias

(1). $\mathrm{E}_\theta(T^*) = \mathrm{E}_\theta(T)$ for any θ,

(2) T^* is a function of S alone and does not involve θ, i.e., T^* is a statistics,

(3). $E_\theta(T^* - \tau(\theta))^2 \leq E_\theta(T - \tau(\theta))^2$, i.e., $MSE_\theta(T^*) \leq MSE_\theta(T)$, for every θ and the equality is strict unless $T = T^*$.

$\because Var(T^*) \leq Var(T)$

(Q: Why T^* better?)

Proof.

1. $\mathrm{E}(T^*) = \mathrm{E}[\mathrm{E}(T|S)] = \mathrm{E}(T)$ by the law of total expectation (LN, CH1~6, p.51)
2. S is sufficient for $\theta \Rightarrow$ distribution of $X_1, \ldots, X_n | S$ not involve θ

   $\Rightarrow$ conditional distribution $T|S$ does not involve θ

   $\Rightarrow T^* = \mathrm{E}(T|S)$ is a function of S only
3. $T|S$

$$\begin{aligned} \mathrm{Var}(T) &= \mathrm{Var}[\mathrm{E}(T|S)] + \mathrm{E}[\mathrm{Var}(T|S)] \\ &\geq \mathrm{Var}[\mathrm{E}(T|S)] = \mathrm{Var}(T^*) \end{aligned}$$

By variance decomposition (LN, CH1~6, p.52)

with equality if and only if $\mathrm{E}[\mathrm{Var}(T|S)] = 0$, i.e. $\mathrm{Var}(T|S) = 0$ with probability 1, which is the case only if T is a function of S, which would imply $T = T^* = E(T|S)$

For all $\underset{\sim}{X}$ s.t. $S(\underset{\sim}{X})=s$, $T(\underset{\sim}{X})$ is a constant

Question 6.9 (uniqueness)

Suppose that two estimates T_1 and T_2 are unbiased (or have same bias). Which of $E(T_1|S)$ and $E(T_2|S)$, where S is sufficient, has smaller variance? Is it possible that there exists only one *unique* function of S that is unbiased (or has same bias)?

Theorem 6.18

If the distribution of S is complete and $\mathrm{E}[u_1(S)] = \tau(\theta)$, $\mathrm{E}[u_2(S)] = \tau(\theta)$ for all θ, then $\mathrm{E}[u_1(S) - u_2(S)] = 0$ for all θ, which implies that $u_1(S) = u_2(S)$ with probability 1 for any θ.

$u_1(S) - u_2(S) = 0,\ \forall S$ (almost sure)

Q. Why can sufficient + complete do the work? (Check graph in LN p.70)

Theorem 6.19 (Lehmann-Scheffe theorem)

Suppose that $X_1, X_2, \ldots, X_n$ have joint pdf/pmf $f(x_1, x_2, \ldots, x_n; \theta)$ and S is a complete and sufficient statistic for θ. If $T^* = T^*(S)$ is a function of S and unbiased for $\tau(\theta)$. Then T^* is the unique UMVUE of $\tau(\theta)$.

unbiased ∵ sufficient

Proof. Suppose T is any other statistic with $\mathrm{E}(T) = \tau(\theta)$ for all θ. $\boxed{\text{Then}}$ $\mathrm{E}(T|S)$ is also unbiased for $\tau(\theta)$ and a function of S. Completeness implies that any statistic that is a function of S and unbiased for $\tau(\theta)$ must be equal to T^* with probability 1. Hence, $T^* = \mathrm{E}(T|S)$ with probability 1 and

$$\mathrm{Var}(\mathrm{E}(T|S)) = \mathrm{Var}(T^*) \leq \mathrm{Var}(T) \text{ for all } \theta.$$

can be any unbiased estimator

Therefore T^* is the unique UMVUE of $\tau(\theta)$.

Summary (methods of finding UMVUE)

1. If $\mathrm{E}(T) = \tau(\theta)$ for all θ and $\mathrm{Var}(T)$ achieves the Cramer-Rao lower bound. Then T is a UMVUE of $\tau(\theta)$. (**Note.** A UMVUE may not attain Cramer-Rao lower bound.)

   Ex 6.32 p.69

2. If S is a complete sufficient statistic, and $u(S)$ satisfies $\mathrm{E}[u(S)] = \tau(\theta)$ for all θ. Then $u(S)$ is a UMVUE of $\tau(\theta)$.

   unbiased

3. If T is an unbiased estimator for $\tau(\theta)$ and S is a complete sufficient statistic for θ. Then $\mathrm{E}(T|S)$ is a UMVUE of $\tau(\theta)$.