7. Asymptotic unbiasedness and consistency; Jan 20, LM 5.7

7.1 Asymptotic unbiasedness LM P406.

Consider estimators based on an n-sample: $T_n = T_n(X_1, ..., X_n)$, where $X_1, ..., X_n$ are i.i.d.

Even estimators that are biased, may be close to unbiased for large n.

Definition: Estimator T_n is said to *asymptotically unbiased* if $b_{T_n}(\theta) \;=\; \mathrm{E}_\theta(T_n) - \theta \;\to\; 0$ as $n \to \infty$.

7.2 Examples

(i) $X_1, ..., X_n$ an n-sample from $U(0, \theta)$; consider estimators based on $W_n = \max_i X_i$.

$\mathrm{E}(W_n) \;=\; n/(n+1)\theta$: $b_{W_n}(\theta) = n\theta/(n+1) - \theta \;=\; -\theta/(n+1) \;\to\; 0$ and $n \to \infty$

W_n is biased, but is asymptotically unbiased: the bias is order $1/n$.

$(n+1)W_n/n$ has expectation θ; it is unbiased for any n.

$(n+2)W_n/(n+1)$ is the estimator with smallest MSE (see 6.5); $\mathrm{E}((n+2)W_n/(n+1)) \;=\; n(n+2)\theta/(n+1)^2$.

$$\text{bias} \;=\; \theta(n(n+2) - (n+1)^2)/(n+1)^2 \;=\; -\theta/(n+1)^2 \;\to\; 0 \text{ and } n \to \infty$$

So this estimator is also asymptotically unbiased: bias is order $1/n^2$.

(ii) $X_1, ..., X_n$ an n-sample from $N(\mu, \sigma^2)$: estimate σ^2 by a multiple of $S^2 = \sum_{i=1}^n (X_i - \overline{X_n})^2$.

The MoM estimator is $T_n = S^2/n$. The unbiased estimator is $S^2/(n-1)$ (see 6.4).

$$b_{T_n}(\sigma^2) \;=\; \mathrm{E}(S^2/n) - \sigma^2 \;=\; (n-1)\sigma^2/n - \sigma^2 \;=\; -\sigma^2/n \;\to\; 0 \text{ and } n \to \infty.$$

7.3 Chebychev inquality LM P.408

The reason we liked estimators with small MSE is that they seemed to give estimators with a probability of being close to the true value of θ. Chebychev's inequalilty makes this relationship explicit.

Chebychev's Inequality: For any random variable W: $P(|W - \theta| > a) \;\le\; \mathrm{E}((W-\theta)^2)/a^2$.

Proof: (Case of continuous W with pdf $f_W(w)$.)

$$\begin{aligned}
\mathrm{E}((W-\theta)^2) &= \int_{-\infty}^{\infty} (w-\theta)^2 f_W(w)\, dw \\
&= \int_{-\infty}^{\theta-a} (w-\theta)^2 f_W(w)\, dw \;+\; \int_{\theta-a}^{\theta+a} (w-\theta)^2 f_W(w)\, dw \;+\; \int_{\theta+a}^{\infty} (w-\theta)^2 f_W(w)\, dw \\
&\ge \int_{-\infty}^{\theta-a} (w-\theta)^2 f_W(w)\, dw \;+\; \int_{\theta+a}^{\infty} (w-\theta)^2 f_W(w)\, dw \\
&\ge a^2 \int_{-\infty}^{\theta-a} f_W(w)\, dw \;+\; a^2 \int_{\theta+a}^{\infty} f_W(w)\, dw \;=\; a^2\, P(|W-\theta| \;>\; a)
\end{aligned}$$

7.4 Consistency LM P.406-7

Consider estimators based on an n-sample: $T_n = T_n(X_1, ..., X_n)$, where $X_1, ..., X_n$ are i.i.d.

Definition: The estimator T_n of θ is *consistent* if, for any $\epsilon > 0$, $P(|T_n - \theta| > \epsilon) \;\to\; 0$ as $n \to \infty$.

But note now from Chebychev's inequlity, the estimator **will** be consistent if $\mathrm{E}((T_n - \theta)^2) \;\to\; 0$ as $n \;\to\; \infty$.

Note also, MSE of T_n is $(b_{T_n}(\theta))^2 \;+\; \mathrm{var}_\theta(T_n)$ (see 5.3).

So the estimator will be consistent if it is asymptotically unbiased, and its variance $\;\to\; 0$ as $n \;\to\; \infty$.

8. Examples of consistency and other properties

8.1 Back to Binomial and Poisson examples

(i) $X_1, ..., X_n$ i.i.d $\sim \mathcal{P}o(\theta)$.

MoM estimator of θ is $T_n = \sum_1^n X_i/n$, and is unbiased $\mathrm{E}(T_n) = \theta$.

Also $\mathrm{var}(T_n) = \theta/n \to 0$ as $n \to \infty$, so the estimator T_n is consistent for θ.

(ii) $X_1, ..., X_n$ i.i.d $\sim Bin(r, \theta)$. MoM estimator of θ is $T_n = \sum_1^n X_i/rn$, and is unbiased $\mathrm{E}(T_n) = \theta$. Also $\mathrm{var}(T_n) = \theta(1-\theta)/rn \to 0$ as $n \to \infty$, so the estimator T_n is consistent for θ.

(Note r is fixed, it is n that $\to \infty$.

8.2 Estimating μ and μ^2

Consider any distribution, with mean μ, and variance σ^2, and $X_1, ..., X_n$ an n-sample from this distribution.

Let $\overline{X_n} = \sum_{i=1}^n X_i/n$. $\mathrm{E}(\overline{X_n}) = \mu$, and $\mathrm{var}(\overline{X_n}) = \sigma^2/n$ (why?).

So $\overline{X_n}$ is unbiased and a consistent estimator of μ. (Why?)

Now suppose we want to estimate μ^2: we could try $\overline{X_n}^2$.

$\mathrm{E}(\overline{X_n}^2) = \mathrm{var}(\overline{X_n}) + (\mathrm{E}(\overline{X_n}))^2 = \sigma^2/n + \mu^2$ (why?)

So $\overline{X_n}^2$ is *not* unbiased for μ^2, but it is asymptotically unbiased.

What about $\mathrm{var}(\overline{X_n}^2) = \mathrm{E}(\overline{X_n}^4) - (\mathrm{E}(\overline{X_n}^2))^2$? In general, this is very messy (so we won't do it), but in fact, provided $\mathrm{E}(X_i^4)$ is finite, $\mathrm{var}(\overline{X_n}^2) \to 0$ as $n \to \infty$.

So in fact (although we have not shown it), $\overline{X_n}^2$ is consistent for μ^2, provided $\mathrm{E}(X_i^4)$ is finite.

8.3 Examples for an n-sample from a uniform $U(0, \theta)$ distrubution

(i) The MoM estimator of θ is $2\overline{X_n} = (2/n)\sum_{i=1}^n X_i$. The estimator has expectation θ and variance $4\mathrm{var}(X_i)/n$, so is unbiased and has variance $\to 0$ as $n \to \infty$. So the estimator is consistent.

(ii) We had also the "better" estimator $(n+1)/n.\max(X_i)$. This was also unbiased and has a smaller variance, in fact of order $1/n^2$. So clearly this one is also consistent.

(iii) What if we just used $W = \max(X_i)$? W has expectation $n\theta/(n+1)$ (so asymptotically unbiased) and also has variance order $1/n^2$. So it is consistent. In fact, we know the cdf $F_W(w) = P(W \le w) = (w/\theta)^n$. so we know $P(|W - \theta| < \epsilon) = P(W > \theta - \epsilon) = 1 - ((\theta - \epsilon)/\theta)^n$. (see LM P.407).

8.4 Careful with the bias

(i) $X_1, ..., X_n$ i.i.d from $f_X(x; \sigma) = (1/2\sigma)\exp(-|x|/\sigma).X_i \sim DE(0, \sigma) \equiv \sigma DE(0, 1)$.

$T_n = \sum_{i=1}^n |X_i|/n$. $\mathrm{var}(T_n) = \mathrm{var}(|X_i|)/n \to 0$ as $n \to \infty$.

$\mathrm{E}(|X_i|) = \int_{-\infty}^{\infty} |x|(1/2\sigma)\exp(-|x|/\sigma)\, dx = \int_0^{\infty}(x/\sigma)\exp(-x/\sigma)\, dx = \sigma$

So T_n is consistent for σ.

(ii) $X_1, ..., X_n$ i.i.d from $N(0, \sigma^2) \equiv \sigma N(0, 1)$.

$T_n = \sum_{i=1}^n |X_i|/n$. $\mathrm{var}(T_n) = \mathrm{var}(|X_i|)/n \to 0$ as $n \to \infty$.

$\mathrm{E}(|X_i|) = \int_{-\infty}^{\infty} |x|(1/(\sqrt{2\pi}\sigma))\exp(-\frac{1}{2}(x/\sigma)^2)\, dx = \sqrt{2/\pi}\int_0^{\infty}(x/\sigma)\exp(-\frac{1}{2}(x/\sigma)^2)\, dx = (\sqrt{2/\pi})\sigma$

The estimator is not aymptotically unbiased, so it cannot be consistent.

9. Moment generating functions LM 3.12

9.1: Definition and basic properties

(i) Definition: $M_X(t) = \mathrm{E}(e^{tX})$, provided expectation exists. Note $M_X(0) \equiv 1$.

Discrete case: $M_X(t) = \sum_x e^{tx} p_X(x)$. Continuous case: $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$.

(ii) Moments: Differentiating: $M_X'(t) = \mathrm{E}(Xe^{tX})$: $M_X'(0) = \mathrm{E}(X)$.

$M_X''(t) = \mathrm{E}(X^2 e^{tX})$, $M_X''(0) = \mathrm{E}(X^2)$. In general: $M_X^{(n)}(0) = \mathrm{E}(X^n)$.

Although this is basis of name "mgf", it is not often useful in practice: there are easier ways!

(iii) Uniqueness: Mgfs are unique. That is, if $M_X(t) = M_Y(t)$ for all t in an open interval containing 0, then X and Y have the same distribution. **This is useful,** as we will see below.

9.2: Examples of mgf's; Discrete (for convenience, write $z = e^t$).

Binomial: $Bin(n,p)$; $q = 1-p$: $\mathrm{E}(z^X) = \sum_{k=0}^{n} \binom{n}{k} (pz)^k q^{n-k} = (q+pz)^n$

Poisson: $\mathcal{P}o(\mu)$: $\mathrm{E}(z^X) = \sum_{k=0}^{\infty} e^{-\mu}(\mu z)^k/k! = \exp(\mu(z-1))$

Geometric: $Geo(p)$: $\mathrm{E}(z^X) = \sum_{k=1}^{\infty} q^{k-1} p z^k = pz/(1-qz)$

Negative binomial: $NegB(r,p)$:

$$\mathrm{E}(z^X) = \sum_{k=r}^{\infty} \binom{k-1}{r-1} q^{k-r} p^r z^k = (pz)^r \sum_{k=0}^{\infty} \binom{k+r-1}{k} (qz)^k = (pz)^r (1-qz)^{-r}$$

9.3: Examples of mgf's: Continuous

Exponential: $\mathcal{E}(\lambda)$: $\mathrm{E}(e^{tX}) = \lambda \int_0^{\infty} \exp(-(\lambda - t)x)\ dx = \lambda/(\lambda - t)$ provided $t < \lambda$.

9.4: Mgf of linear functions and sums of independent r.vs

(i) Let $Y = aX + b$: $M_Y(t) = \mathrm{E}(\exp((aX+b)t)) = e^{bt}\mathrm{E}(\exp((at)X)) = e^{bt} M_X(at)$.

(ii) Let $Y = aX$ where X is exponental $\mathcal{E}(\lambda)$, then

$M_Y(t) = M_X(at) = \lambda/(\lambda - at) = (\lambda/a)/((\lambda/a) - t)$ which is the Mgf of $\mathcal{E}(\lambda/a)$.

So, by uniqueness of Mgf, aX is distributed as $\mathcal{E}(\lambda/a)$.

(iii) Let X and Y be independent random variables: $W = X + Y$:

$M_W(t) = \mathrm{E}(\exp((X+Y)t) = \mathrm{E}(\exp(Xt)\exp(Yt)) = \mathrm{E}(e^{Xt})\mathrm{E}(e^{Yt}) = M_X(t) M_Y(t)$

(iv) Let $X_1, ..., X_n$ be i.i.d. with same dsn as X. $W = \sum_{i=1}^{n} X_i$

$M_W(t) = \prod_{i=1}^{n} M_{X_i}(t) = (M_X(t))^n$.

9.5 Immediate conclusions!!

Sum of independent Binomials (same p) is Binomial;

Sum of independent Poisson (any means) is Poisson

Sum of independent Geometrics (same p) is Negative Binomial; and of NegBin is also NegBin.