Statistics 581
Problem Set 5 Solutions
Wellner; 11/6/2002

1. Suppose that $X_i \sim \text{Bernoulli}(p_i)$, $i = 1, \ldots, n$ are independent. Show that if

$$\sum_{i=1}^{n} p_i(1-p_i) \to \infty\,, \tag{0.1}$$

then

$$\frac{\sqrt{n}(\overline{X}_n - \overline{p}_n)}{\sqrt{n^{-1}\sum_{i=1}^{n} p_i(1-p_i)}} \to_d N(0,1)\,.$$

Give one example $\{p_i\}_{i\ge 1}$ for which (0.1) holds and another example for which it fails.

Solution: With $X_{ni} \equiv X_i - p_i$, $i = 1, \ldots, n, \ldots$ we have $E(X_{ni}) = 0$, $\sigma_{ni}^2 = Var(X_{ni}) = p_i(1-p_i)$, and

$$\begin{aligned} \gamma_{ni} &= E|X_{ni}|^3 = E|X_i - p_i|^3 = |1-p_i|^3 p_i + |0-p_i|^3(1-p_i) \\ &\le p_i(1-p_i)\{(1-p_i)^2 + p_i^2\} \le 2p_i(1-p_i)\,, \end{aligned}$$

so that $\sigma_n^2 = \sum_{i=1}^{n} p_i(1-p_i)$ and $\gamma_n \le 2\sum_{i=1}^{n} p_i(1-p_i)$. hence

$$\frac{\gamma_n}{\sigma_n^3} \le \frac{2}{\{\sum_{i=1}^{n} p_i(1-p_i)\}^{1/2}} \to 0$$

if $\sum_1^n p_i(1-p_i) \to \infty$. Hence it follows from the Liapunov CLT that

$$\frac{\sum_{i=1}^{n}(X_i - p_i)}{\sqrt{\sum_1^n p_i(1-p_i)}} \to_d N(0,1)\,,$$

and this is equivalent to the stated conclusion. Note that this generalizes the result of Problem #4, Problem Set 3.
If $p_i = 1/i^r$ with $r > 1$, then the assumption fails:

$$\sum_{i=1}^{n} p_i(1-p_i) = \sum_{i=1}^{n} i^{-r} - \sum_{i=1}^{n} i^{-2r} \to \sum_{i=1}^{\infty} i^{-r} - \sum_{i=1}^{\infty} i^{-2r} < \infty\,.$$

On the other hand, if $p_i = 1/i$, then it holds:

$$\sum_{i=1}^{n} p_i(1-p_i) = \sum_{i=1}^{n} i^{-1} - \sum_{i=1}^{n} i^{-2} \to \infty - \sum_{i=1}^{\infty} i^{-2} = \infty\,.$$

2. Suppose that $X_1, \ldots, X_n$ are independent with common mean μ, but with variances $\sigma_1^2, \ldots, \sigma_n^2$ respectively.
(a) Show that $\overline{X}_n$ is a consistent estimator of μ if $\sum_{i=1}^n \sigma_i^2 = o(n^2)$.
(b) Now suppose that $X_i = \mu + \sigma_i \epsilon_i$ where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with some distribution function F with $E(\epsilon_1) = 0$ and $Var(\epsilon_1) = 1 < \infty$. Show that if

$$\max_{1\le i\le n} \sigma_i^2 / \sum_{i=1}^{n} \sigma_i^2 \to 0 \tag{0.2}$$

then with $\overline{\sigma}_n^2 \equiv n^{-1}\sum_1^n \sigma_i^2$,

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\overline{\sigma}_n} \to_d N(0,1)\,. \tag{0.3}$$

Hence show that if both (0.2) and

$$\overline{\sigma_n^2} \to \text{"something"} \equiv \sigma_0^2\,, \tag{0.4}$$

then

$$\sqrt{n}(\overline{X}_n - \mu) \to_d N(0, \sigma_0^2)\,.$$

(c) Show that (0.2) holds but that (0.4) fails if $\sigma_i^2 = Ai^r$ with $r < 1$. Hence show that in this case $n^{(1-r)/2}(\overline{X}_n - \mu) = O_p(1)$.

Solutions: (a) Let $\epsilon > 0$. Note that

$$P(|\overline{X}_n - \mu| > \epsilon) \le \frac{Var(\overline{X}_n)}{\epsilon^2} = \frac{\sum_{i=1}^n \sigma_i^2}{n^2\epsilon^2} \to 0$$

if $\sum_1^n \sigma_i^2 = o(n^2)$.
(b) For clarity, change notation by letting $\sigma_i \equiv a_i$. Set $X_{ni} \equiv X_i - \mu = a_i\epsilon_i$ for $i = 1, \ldots, n$. Then $E(X_{ni}) = 0$, $\sigma_{ni}^2 = Var(X_{ni}) = a_i^2$, and $\sigma_n^2 = \sum_{i=1}^n a_i^2$. To check the Lindeberg condition we compute

$$\begin{aligned}\frac{1}{\sigma_n^2}\sum_{i=1}^{n} E|X_{ni}|^2 1_{[|X_{ni}|>\delta\sigma_n]} &= \frac{1}{\sigma_n^2}\sum_{i=1}^{n} E\{a_i^2\epsilon_i^2 1_{a_i|\epsilon_i|\ge\delta\sigma_n]}\} \\ &\le E\{\epsilon_1^2 1_{[|\epsilon_1|>\delta\sigma_n/\max_{1\le i\le n} a_i]}\} \to 0\end{aligned}$$

if (0.2) holds. Thus (0.3) follows from the Lindeberg-Feller CLT.

(c) If $a_i^2 = Ai^r$, then

$$\sum_{i=1}^n a_i^2 = A \sum_{i=1}^n i^r \sim \frac{An^{r+1}}{r+1} \qquad \text{as } \ n \to \infty \,,$$

(since $n^{-1} \sum_1^n (i/n)^r \to \int_0^1 x^r dx = 1/(r+1)$). Thus

$$\frac{\max_{i \le n} a_i^2}{\sum_1^n a_i^2} = \frac{On^r}{n^{r+1}} \to 0 \,,$$

but

$$\frac{1}{n} \sum_{i=1}^n \sigma_i^2 \sim \frac{An^r}{r+1} \to \infty \,.$$

3. Suppose that $X_1, \ldots, X_n$ are independent with common mean μ, but with variances $\sigma_1^2, \ldots, \sigma_n^2$ respectively, exactly as in problem 2 above. Consider estimators of μ of the form $T_n \equiv T_n(w) = \sum_{i=1}^n w_{ni} X_i$ where $w = w_n = (w_{n1}, \ldots, w_{nn})$ is a vector of weights with $\sum_{i=1}^n w_{ni} = 1$.
(a) Show that all the estimators $T_n(w)$ are unbiased, and that the choice of weights which minimizes $Var(T_n(w))$ is

$$(0.5) \qquad w_{ni}^{opt} = \frac{1/\sigma_i^2}{\sum_{j=1}^n (1/\sigma_j^2)} \qquad \text{for } \ i = 1, \ldots, n \,.$$

(b) Compute $Var(T_n(w^{opt}))$ and show that $T_n(w^{opt})$ is a consistent estimator of μ if $\sum_1^n (1/\sigma_j^2) \to \infty$.
(c) Now suppose that $X_i = \mu + \sigma_i \epsilon_i$ where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with some distribution function F with $E(\epsilon_1) = 0$ and $Var(\epsilon_1) = 1 < \infty$ as in 2(b) above. Show that

$$\sqrt{\sum_1^n (1/\sigma_i^2)} \left(T_n(w^{opt}) - \mu \right) \to_d N(0,1)$$

if

$$\frac{\max_{1 \le i \le n} (1/\sigma_i^2)}{\sum_{j=1}^n (1/\sigma_j^2)} \to 0 \,.$$

(d) Compute $Var[T_n(w^{opt})]/Var[\overline{X}_n]$ in the case $\sigma_i^2 = Ai^r$ for $r = .25,\ .50,\ .75,\ 1$ and $n = 5, 10, 20, 50, 100$, and ∞.

Solution: (a) Unbiasedness of the estimators $T_n(w)$ is trivial:

$$E(T_n(w)) = \sum_{i=1}^n w_{ni} E(X_i) = \sum_{i=1}^n w_{ni}\mu = \mu \sum_{i=1}^n w_{ni} = \mu\,.$$

By direct calculation,

$$Var(T_n(w)) \;=\; \sum_{i=1}^n w_{ni}^2 \sigma_i^2$$

and we want to minimize this subject to $\sum_{i=1}^n w_{ni} = 1$. Thus define

$$V^2(w,\lambda) = \sum_{i=1}^n w_{ni}^2\sigma_i^2 + \lambda(\sum_{i=1}^n w_{ni} - 1)\,.$$

Thus

$$\frac{\partial}{\partial w_{ni}} V^2(w,\lambda) = 2w_{ni}\sigma_i^2 + \lambda, \qquad i = 1, \ldots, n\,,$$

and setting this equal to 0 for all i yields

$$w_{ni} = -\frac{\lambda/2}{\sigma_i^2}$$

where, in order to satisfy the constraint,

$$1 = \sum_{i=1}^n w_{ni} = -\frac{\lambda}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2}$$

and hence

$$\lambda = \frac{-2}{\sum_{i=1}^n (1/\sigma_i^2)}\,.$$

Thus the optimal choice of the weights w_{ni} is given by

$$w_{ni}^{opt} = \frac{1/\sigma_i^2}{\sum_{j=1}^n (1/\sigma_j^2)}\,, \qquad i = 1, \ldots, n\,.$$

Here is a more geometric proof involving a "dual" optimization problem. Write $w_i \equiv w_{ni}$. We want to minimize

$$\sum_{i=1}^{n} w_i^2 \sigma_i^2 = w^T \text{diag}(\sigma_i^2) w$$

subject to the constraint $\sum_1^n w_i = w^T 1 = 1$. Alternatively, letting $v_i \equiv w_i \sigma_i$, we want to minimize

$$v^T v \qquad \text{subject to} \qquad \langle v, 1/\sigma \rangle = 1$$

where $1/\sigma = (1/\sigma_1, \ldots, 1/\sigma_n)^T$. Geometrically this amounts to finding the radius of the smallest ball which intesects the hyperplane determined by the vector $1/\sigma$ and the magnitude constant 1. Now the dual problem is: maximize

$$\langle v, 1/\sigma \rangle \qquad \text{subject to} \qquad v^T v \leq C^2 \, .$$

Geometrically this amounts to finding the hyperplane with the largest magnitude constant which intersects the ball with radius C. But the Cauchy-Schwarz inequality,

$$\langle v, 1/\sigma \rangle \leq \sqrt{v^T v (1/\sigma)^T (1/\sigma)} \leq C \sqrt{\sum_{i=1}^{n} (1/\sigma_i^2)} = 1 \tag{0.6}$$

if $C = 1/\sqrt{\sum_{i=1}^n (1/\sigma_i^2)}$, and equality holds in the first inequality of (0.6) if $v = A(1/\sigma)$ for some A. Translating back to w gives $w_i = w_{ni}$ as in the first solution.

(b) The resulting minimal variance is

$$Var[T_n(w^{opt})] = \sum_{i=1}^{n} [w_{ni}^{opt}]^2 \sigma_i^2 = \frac{\sum_{i=1}^n (1/\sigma_i^2)}{(\sum_{i=1}^n (1/\sigma_i^2))^2} = \frac{1}{\sum_{i=1}^n (1/\sigma_i^2)} \, .$$

Hence by Chebychev's inequality

$$P(|T_n(w^{opt}) - \mu| > \epsilon) \leq \frac{Var[T_n(w^{opt})]}{\epsilon^2} = \frac{1}{\epsilon^2 \sum_{i=1}^n (1/\sigma_i^2)} \to 0$$

for every $\epsilon > 0$ if $\sum_1^n (1/\sigma_i^2) \to \infty$.
(c) When $X_i = \mu + \sigma_i \epsilon_i$, and we want to show that

$$\sqrt{\sum_1^n (1/\sigma_i^2)}\,\big(T_n(w^{opt}) - \mu\big) \to_d N(0,1)\,,$$

it is convenient (for clarity in avoiding clashes with the notation of the Lindeberg-Feller CLT) to temporarily relabel the σ_i as a_i, $i = 1, \ldots, n$. Note that then the quantity on the left side in the last display becomes

$$\sqrt{\sum_1^n (1/a_i^2)}\,\big(T_n(w^{opt}) - \mu\big) = \frac{\sum_{i=1}^n (1/a_i^2) a_i \epsilon_i}{\sqrt{\sum_{i=1}^n (1/a_i^2)}} = \sum_{i=1}^n c_{ni} \epsilon_i$$

where

$$c_{ni} = \frac{1/a_i}{\sqrt{\sum_{i=1}^n (1/a_i^2)}}\,, \qquad i = 1, \ldots, n\,.$$

By the same argument we have used several times now for weight sums of i.i.d. random variables with mean zero and finite variance, the Lindeberg condition holds if

$$\max_{1 \le i \le n} |c_{ni}| = \frac{max_{1 \le i \le n}(1/a_i)}{\sqrt{\sum_{i=1}^n (1/a_i^2)}} \to 0\,,$$

and we conclude that (0.6) holds.
(d) When $\sigma_i^2 = A i^r$ for $i = 1, \ldots, n$ We have

$$Var(\overline{X}_n) = \frac{\sum_{i=1}^n \sigma_i^2}{n^2} = \frac{A \sum_{i=1}^n i^r}{n^2}\,,$$

while

$$Var[T_n(w^{opt})] = \frac{1}{\sum_{i=1}^n (1/\sigma_i^2)} = \frac{A}{\sum_{i=1}^n i^{-r}}\,.$$

Hence

$$\frac{Var[T_n(w^{opt})]}{Var[\overline{X}_n]} = \frac{n^2}{\left(\sum_{i=1}^n i^{-r}\right)\left(\sum_{i=1}^n i^r\right)}\,.$$

Computing this for $r = .25,\ .50,\ .75,\ 1$ and $n = 5, 10, 20, 50, 100,$ and ∞, and noting that the above ratio equals

$$\begin{aligned}&\frac{1}{\left(n^{-1}\sum_1^n (i/n)^{-r}\right)\left(n^{-1}\sum_1^n (i/n)^{r}\right)}\\ &\to \frac{1}{\int_0^1 x^{-r}dx \int_0^1 x^r dx} = (1-r)(1+r) = 1 - r^2\end{aligned}$$

(where the convergence holds for $0 < r \le 1$ even though the first integral in the denominator is infinite for $r = 1$), yields the following table.

Table 1: Efficiencies of $\overline{X}_n$ relative to $T_n(w^{opt})$

n	5	10	20	50	100	∞
$r = .25$	.980	.970	.961	.952	.948	.9375
$r = .50$	.923	.886	.854	.820	.801	.7500
$r = .75$	.836	.763	.700	.633	.595	.4375
$r = 1.0$	.730	.621	.529	.436	.382	0

4. Suppose that $X_1, \ldots, X_n$ are i.i.d. random vectors with values in R^k with $E(X_1) = \mu$ and $E(X_1^T X_1) < \infty$ so that $\Sigma = E(X_1 - \mu)(X_1 - \mu)^T$ is well-defined. Thus

$$Z_n \equiv \sqrt{n}(\overline{X}_n - \mu) \to_d Z \sim N_k(0, \Sigma)\,.$$

Suppose that $g : R^k \to R$ is a function, and suppose that $\nabla g = \dot{g}$ exists at μ. Then the delta-method (or g' theorem) tells us that

$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \to_d \nabla g(\mu)^T Z \sim N(0, \nabla g(\mu)^T \Sigma \nabla g(\mu))\,.$$

Show that we can strengthen this as follows: Suppose that $\nabla g = \dot{g}$ is continuous at μ. Then $\sqrt{n}(g(\overline{X}_n) - g(\mu))$ is *asymptotically linear* at μ:

$$\begin{aligned}\sqrt{n}(g(\overline{X}_n) - g(\mu)) &= \nabla g(\mu)^T \sqrt{n}(\overline{X}_n - \mu) + o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i) + o_p(1)\end{aligned}$$

where

$$\psi(x) = \nabla g(\mu)^T (x - \mu) \tag{0.7}$$

which is called the *influence function* of $g(\overline{X}_n)$ as an estimator of $g(\mu)$, has mean $E\psi(X_i) = 0$ and $Var(\psi(X_i)) = \nabla g(\mu)^T \Sigma \nabla g(\mu)$.

Solution: By Taylor's theorem, for some Y_n^* satifying $|Y_n^* - \mu| \leq |\overline{X}_n - \mu| \to_p 0$ it follows that

$$\begin{aligned}
\sqrt{n}(g(\overline{X}_n) - g(\mu)) &= \nabla g(Y_n^*)\sqrt{n}(\overline{X}_n - \mu) \\
&= \nabla g(\mu)\sqrt{n}(\overline{X}_n - \mu) \\
&\qquad + \{\nabla g(Y_n^*) - \nabla g(\mu)\}\sqrt{n}(\overline{X}_n - \mu) \\
&= \nabla g(\mu)\sqrt{n}(\overline{X}_n - \mu) + o_p(1)
\end{aligned}$$

since $\nabla g(Y_n^*) \to_p \nabla g(\mu)$ by continuity of ∇g at μ and since $\sqrt{n}(\overline{X}_n - \mu) = O_p(1)$. Now note that

$$\nabla g(\mu)\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \nabla g(\mu)(X_i - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(X_i)$$

with ψ as in (0.7).