Statistical Theory MATH-442
Spring 2025 Ho Yun & Yoav Zemel

Exercise sheet 2

You may wish to refer to the law of large numbers and the central limit theorem on the slides of week one.

Exercise 1 Let X_n, X be random variables taking values in $\mathbb{Z}$ and such that $\mathbb{P}(X_n = k) \to \mathbb{P}(X = k)$ for all $k \in \mathbb{Z}$. Show that $X_n \xrightarrow{d} X$.

Exercise 2 Give a counterexample to show that neither of $X_n \xrightarrow{p} X$ or $X_n \xrightarrow{d} X$ ensures that $\mathrm{E}X_n \to \mathrm{E}X$ as $n \to \infty$.

Solution 1 For all $a \le b \in \mathbb{R}$ we have

$$P(a < X_n \le b) = \sum_{k=\lceil a \rceil^+}^{\lfloor b \rfloor} P(X_n = k) \to \sum_{k=\lceil a \rceil^+}^{\lfloor b \rfloor} P(X = k) = P(a < X \le b)$$

where $\lceil a \rceil^+$ is the smallest integer strictly larger than a. Therefore the distribution functions satisfy

$$F_{X_n}(b) - F_{X_n}(a) \to F_X(b) - F_X(a), \qquad a \le b,$$

and so for all $a, b \in \mathbb{R}$

$$\limsup_n F_{X_n}(b) \le F_X(b) - F_X(a) \tag{1}$$

$$\liminf_n F_{X_n}(a) \ge 1 - F_X(b) + F_X(a) \tag{2}$$

Taking $a \to -\infty$ in (1) proves $\limsup_n F_{X_n}(b) \le F_X(b)$ for all $b \in \mathbb{R}$. Taking $b \to \infty$ in (2) yields $\liminf_n F_{x_n}(a) \ge F_X(a)$ for all $a \in \mathbb{R}$. Together this gives $F_{X_n}(x) \to F_X(x)$ for all $x \in \mathbb{R}$, which implies convergence in distribution.

Solution 2 Let X_n satisfy $P(X_n = n) = 1/(n+1)$, $P(X_n = 1/n) = n/(n+1)$. Then $X_n \xrightarrow{p} 0$ and $X_n \xrightarrow{d} 0$ but $\mathrm{E}X_n = 1 \to 1 \ne \mathrm{E}0$.

Exercise 3

Find the limit in distribution (as $n \to \infty$) for the sequence $\{X_n\}_{n\in\mathcal{N}}$ defined as:

(a) $\{E_k\}_{k\in\mathcal{N}}$ iid, $E_k \sim \text{Exp}(1)$ for every $k \in \mathcal{N}$,

$$X_n = \tfrac{1}{\sqrt{n}} \sum_{k=1}^{n} E_k - \sqrt{n},$$

(b) $\{U_k\}_{k\in\mathcal{N}}$ iid, U_k uniform on $(0,1)$ for every $k \in \mathcal{N}$,

$$X_n = \min(U_1, \dots U_n),$$

(c) $\{U_k\}_{k\in\mathcal{N}}$ iid, U_k uniform on $(0,1)$ for every $k \in \mathcal{N}$,

$$X_n = n \times \min(U_1, \dots U_n),$$

(d) $\{U_k\}_{k\in\mathcal{N}}$ iid, U_k uniform on $(0,1)$ for every $k \in \mathcal{N}$,

$$X_n = \sqrt{n} \times \min(U_1, \dots U_n),$$

(e) $X_n \sim \text{B}(n, p_n)$ such that $\lim_{n\to\infty} p_n = 0$ and $\lim_{n\to\infty} np_n = \lambda$.

Exercise 4 Prove that if $\text{E}X_n \to a$ (a is a finite constant) and $\text{var}X_n \to 0$, then $X_n \xrightarrow{p} a$ as $n \to \infty$.

Solution 3

(a) $X \sim \text{N}(0,1)$,

(b) $X = 0$ in probability (one can prove that the limit hold a.s. using Borel-Cantelli's lemma),

(c) $X \sim \text{Exp}(1)$,

(d) $X = 0$ in probability (one can prove that the limit hold a.s. using Borel-Cantelli's lemma),

(e) $X \sim \text{Poisson}(\lambda)$.

Solution 4 $P(|X_n - a| > \epsilon) \le P(|X_n - \text{E}X_n| > \epsilon/2) + P(|\text{E}X_n - a| > \epsilon/2) \le \frac{4}{\epsilon^2}\text{var}X_n + P(|\text{E}X_n - a| > \epsilon/2) \to 0$ as $n \to \infty$.

Exercise 5 Let $h : [0,1] \to [0,1]$ be a continuous function. We are interested in computing its integral $\int_0^1 h(t)dt$ by Monte Carlo simulation.

(i) Let $\xi_1, \xi_2, \ldots, \eta_1, \eta_2, \ldots$ be independent random variables uniformly distributed on $[0,1]$ and let $X_k = 1_{[\eta_k \le h(\xi_k)]}$. Show that $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ converges almost surely to $\int_0^1 h(t)dt$.

(ii) Let $\xi_1, \xi_2, \ldots$ be independent random variables uniformly distributed on $[0,1]$ and let $Y_k = h(\xi_k)$. Show that $\overline{Y}_n = \frac{1}{n}\sum_{k=1}^n Y_k$ converges almost surely to $\int_0^1 h(t)dt$.

(iii) Compute and compare $\mathrm{var}\overline{X}_n$ and $\mathrm{var}\overline{Y}_n$.

Solution 5

(i) Compute $\mathrm{E}X_1 = \mathrm{E}1_{[\eta_1 \le h(\xi_1)]} = P(\eta_1 \le h(\xi_1)) = \int_{\{(\xi,\eta):\eta \le h(\xi)\}} d\xi\, d\eta = \int_0^1 h(t)dt$, or alternatively $\mathrm{E}X_1 = \mathrm{E}1_{[\eta_1 \le h(\xi_1)]} = P(\eta_1 \le h(\xi_1)) = \mathrm{E}P(\eta_1 \le h(\xi_1)|\xi_1) = \mathrm{E}h(\xi_1) = \int_0^1 h(t)dt$. Therefore, by the Strong Law of Large Numbers, $\overline{X}_n \xrightarrow{a.s.} \int_0^1 h(t)dt$.

(ii) Similarly, compute $\mathrm{E}Y_1 = \mathrm{E}h(\xi_1) = \int_0^1 h(t)dt$ and conlude that by the Strong Law of Large Numbers, $\overline{Y}_n \xrightarrow{a.s.} \int_0^1 h(t)dt$.

(iii) The variable X_1 is Bernoulli distributed with success probability $\mathrm{E}X_1 = \int_0^1 h(t)dt$, thus its variance is $\left(\int_0^1 h(t)dt\right)\left(1 - \int_0^1 h(t)dt\right)$. The variance of Y_1 is $\mathrm{E}Y_1^2 - (\mathrm{E}Y_1)^2 = \int_0^1 h(t)^2 dt - \left(\int_0^1 h(t)dt\right)^2$. Now $\mathrm{var}\overline{X}_n - \mathrm{var}\overline{Y}_n = \frac{1}{n}\int_0^1 (h(t) - h(t)^2)dt \ge 0$. Hence $\overline{Y}_n$ is more accurate as an estimator of $\int_0^1 h(t)dt$.

Exercise 6

(a) Let $X_1, \dots, X_n$ be a sample of Bernoulli distributed variables with success probability p. We are interested in estimating the odds defined as $r = \frac{p}{1-p}$. The sample mean $\widehat{p} = n^{-1} \sum_{i=1}^n X_i$ is a meaningful estimator of $p = \mathrm{E}X$. Then r is naturally estimated by $\widehat{r} = \frac{\widehat{p}}{1-\widehat{p}}$. Find the asymptotic distribution of $\widehat{r}$, that is, investigate the convergence in distribution of $n^{1/2}(\widehat{r} - r)$ as $n \to \infty$.

(b) Let $X_1, \dots, X_n$ be a sample from a Poisson distribution with intensity $\lambda > 0$. We might be interested in estimating $\pi = P(X = 1) = \lambda e^{-\lambda}$. The sample mean $\widehat{\lambda} = n^{-1} \sum_{i=1}^n X_i$ is a meaningful estimator of $\lambda = \mathrm{E}X$. Then π is naturally estimated by $\widehat{\pi} = \widehat{\lambda} e^{-\widehat{\lambda}}$. Find the asymptotic distribution of $\widehat{\pi}$, that is, investigate the convergence in distribution of $n^{1/2}(\widehat{\pi} - \pi)$.

(c) Let $X_1, \dots, X_n$ be a sample from a geometric distribution with success probability $p \in (0, 1)$ (i.e., $P(X_i = k) = (1-p)^k p$, $k = 0, 1, \dots$). We might be interested in estimating $\pi = P(X > k) = (1-p)^{k+1}$, $k = 0, 1, \dots$ The sample mean $\widehat{\mu} = n^{-1} \sum_{i=1}^n X_i$ is a meaningful estimator of $\mu = \mathrm{E}X = (1-p)/p$. Then $p = 1/(\mu + 1)$ could be estimated by $1/(\widehat{\mu} + 1)$, and thus $\pi = [\mu/(\mu + 1)]^{k+1}$ by $\widehat{\pi} = [\widehat{\mu}/(\widehat{\mu} + 1)]^{k+1}$. Find the asymptotic distribution of $\widehat{\pi}$, that is, investigate the convergence in distribution of $n^{1/2}(\widehat{\pi} - \pi)$. (Recall that $\mathrm{var} X_i = (1-p)/p^2 = \mu(\mu + 1)$.)

Solution 6

(a) $n^{1/2}(\widehat{p} - p) \xrightarrow{d} \mathrm{N}(0, p(1-p))$ by CLT, $r = g(p) = p/(1-p)$, $g'(\mu) = 1/(1-p)^2$, $n^{1/2}(\widehat{r} - r) \xrightarrow{d} 1/(1-p)^2 \mathrm{N}(0, p(1-p)) = \mathrm{N}(0, p/(1-p)^3)$ by the delta method.

(b) $n^{1/2}(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathrm{N}(0, \lambda)$ by CLT, $\pi = g(\lambda) = \lambda e^{-\lambda}$, $g'(\lambda) = (1-\lambda)e^{-\lambda}$, $n^{1/2}(\widehat{\pi} - \pi) \xrightarrow{d} (1-\lambda)e^{-\lambda} \mathrm{N}(0, \lambda) = \mathrm{N}(0, \lambda(1-\lambda)^2 e^{-2\lambda})$ by the delta method.

(c) $n^{1/2}(\widehat{\mu} - \mu) \xrightarrow{d} \mathrm{N}(0, \mu(\mu+1))$ by CLT, $\pi = g(\mu) = [\mu/(\mu+1)]^{k+1}$, $g'(\mu) = (k+1)\mu^k/(\mu+1)^{k+2}$, thus

$$n^{1/2}(\widehat{\pi} - \pi) \xrightarrow{d} (k+1)\tfrac{\mu^k}{(\mu+1)^{k+2}} \mathrm{N}(0, \mu(\mu+1)) = \mathrm{N}\left(0, (k+1)^2 \tfrac{\mu^{2k+1}}{(\mu+1)^{2k+3}}\right)$$

by the delta method.

Exercise 7 Let $X_1, X_2, \ldots$ be a sequence of independent variables with distribution

$$P[X_n = 1] = P[X_n = -1] = \tfrac{1}{2} - \tfrac{1}{2^{n+1}}, \qquad P[X_n = 2^n] = P[X_n = -2^n] = \tfrac{1}{2^{n+1}}.$$

Prove that $n^{-1} \sum_{i=1}^n X_i \to 0$ almost surely as $n \to \infty$.
Hint: The variables X_n are not identically distributed, hence the strong law of large numbers for iid variables does not apply here. But the sequence $\{X_n\}$ is close to some sequence $\{Y_n\}$ which is an iid sequence and satisfies the SLLN. Each variable Y_n is obtained from X_n by setting Y_n to X_n where X_n is ± 1 and setting it to ± 1 where X_n has 'problematic' values $\pm 2^n$. To justify this approximation (the negligibility of the modification from X_n to Y_n), use the Borel–Cantelli lemma:

> **Borel–Cantelli lemma.** Let $A_1, A_2, \ldots$ be a sequence of events. If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then
>
> $$P(A_n \text{ happens for infinitely many values of } n) = 0.$$

Solution 7 Defining Y_n as in the hint and letting $A_k = \{X_k \neq Y_k\}$, we have $\sum P(A_k) = \sum 2^{-k-1} < \infty$, so almost surely A_k happens finitely many times, and in that case $\sum_{i=1}^n (Y_n - X_n)/n \to 0$ since $Y_n = X_n$ for n large, and the factor $1/n$ together with $|Y_n - X_n| \le 2^{\max\{k : A_k \text{ happens}\}} < \infty$ deals with the finitely many potentially problematic cases. So with probability one, in obvious notation

$$\overline{X}_n = \overline{Y}_n + (\overline{X}_n - \overline{Y}_n) \xrightarrow{a.s.} \mathbb{E}Y_1 + 0 = 0,$$

by the strong law of large numbers for the iid sequence (Y_n).