stats2201sampling 625.2 - Samples, distributions and convergence

Gunnar Stefansson

23. september 2021

Efnisyfirlit

1 Sampling, distributions and convergence

1.1 Random sample

1.1.1 Definition of a random sample

A random sample is a collection of random variables which are independent and identically distributed (i.i.d.).

1.1.2 Handout

Definition 1.1. A collection of random variables $X_1, \ldots, X_n$ are a random sample if they are independent and identically distributed.

Typical usage: "let $X_1, \ldots, X_n$ be i.i.d.ǫr "let $X_1, \ldots, X_n \sim f_\theta$ be i.i.d.ǫr "let $X_1, \ldots, X_n \sim F$ be i.i.d.ǫr "let $X_1, \ldots, X_n$ be i.i.d. n(0,1)".
In this type of usage, $\{f_\theta\}$ refers to a family indexed by the unknown parameter θ and F is a cumulative distribution function (c.d.f.).

1.2 Convergence concepts and Chebychev's theorem

1.2.1 Handout

Convergence concepts

Theorem 1.1 (Chebychev or Markov's inequality) Let X be a continuous random variable and $g \geq 0$ be a continuous function. Then for $r \geq 0$:

$$P[g(X) \geq r] \leq \frac{\mathbb{E}[g(X)]}{r}.$$

Sönnun.

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_{-\infty}^{+\infty} g(x) f(x)\, dx \qquad [f \text{ is the density of } X] \\
&= \int_{\{x:g(x)<r\}} g(x) f(x)\, dx + \int_{\{x:g(x)\geq r\}} g(x) f(x)\, dx \\
&\geq \int_{\{x:g(x)\geq r\}} g(x) f(x)\, dx \qquad [g \geq 0] \\
&\geq \int_{\{x:g(x)\geq r\}} r f(x)\, dx = r \int_{\{x:g(x)\geq r\}} f(x)\, dx \\
&= rP[g(X) \geq r]
\end{aligned}$$

Where the integral over $\{x : g(x) \geq r\}$ is well defined since $\{x : g(x) \geq r\} = g^{-1}(]-\infty, r[)$ and g is continuous. Similarly for $\{x : g(x) < r\}$. $\square$

Definition 1.2. A sequence of random variables $X_1, \dots,$ converges to the random variable X in probability if $P[|X_n - X| < \varepsilon] \xrightarrow[n\to\infty]{} 1$ is true for all $\varepsilon > 0$. We write $X_n \xrightarrow{P} X$.

Theorem 1.2 (weak law of large numbers) If $X_1, X_2, \dots$ are independent and identically distributed (iid) random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{V}X_i = \sigma^2 < \infty$, then:

$$\bar{X}_n \xrightarrow{P} \mu,$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Sönnun. $P[|\bar{X}_n - \mu| > \varepsilon] \le \frac{\sigma^2/n}{\varepsilon^2} \xrightarrow[n\to\infty]{} 0$ (from the Chebychev inequality). □

1.3 Estimators

1.3.1 Handout

Definition 1.3. An *estimator* is a (measurable) function of random variables $X_1, \dots, X_n$. Commonly "an estimator" is of the form $T_n = h(X_1, \dots, X_n)$, where $X_1, X_2, \dots$ is a sequence of random variables, i.e. term "the estimator" actually refers to a sequence of estimators.
An estimator T is said to be *unbiased* for a parameter θ if $\mathrm{E}T_n = \theta$. An estimator T_n is said to be *consistent* for θ if $T_n \xrightarrow{P} \theta$.

Example 1.1. If $X_1, X_2, \dots$ are i.i.d. and $\mathrm{E}X_i^4 < \infty$, then

$$S_n^2 \xrightarrow{p} \sigma^2,$$

where

$$S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$$

This is true since

$$P[|S_n^2 - \sigma^2| \ge \varepsilon] \le \frac{V[S_n^2]}{\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

if $V[S_n^2] \to 0$, which holds since

$$V[S^2] = \frac{1}{n}\left(\Theta_4 - \frac{n-3}{n-1}\Theta_2^2\right) \to 0$$

(see e.g. example in Casella and Berger.)
Recall that if the variables are also Gaussian, then

$$W_n := \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

so that

$$V[W_2] = 2(n-1)$$

and

$$V[S^2] = V\left[\frac{\sigma^2}{n-1}W\right] = \frac{\sigma^4}{(n-1)^2} \cdot V[W] = \frac{\sigma^4}{(n-1)^2}2(n-1) = \frac{2\sigma^4}{n-1} \to 0.$$

Theorem 1.3 If $X_n \xrightarrow{P} X$ and h is a continuous function, then $h(X_n) \xrightarrow{P} h(X)$.

The proof is left to the reader (use the definition of continuity).

Example 1.2. Toss a biased coin n times with independent tosses to obtain the random variables $X_n \sim b(n,p)$. Define $\hat{p}_n := \frac{X_n}{n}$. This will have the same distribution as $\bar{Y}_n$ where $Y_1, Y_2, \ldots$ are the outomes of individual tosses and $Y_1, Y_2, \ldots$ are i.i.d. Thus we have

$$\hat{p}_n \xrightarrow{P} p,$$

i.e. $P[|\hat{p}_n - p| > \varepsilon] \xrightarrow[n\to\infty]{} 0$ for all $\varepsilon > 0$.

Example 1.3. $X_n : \underbrace{[0,1]}_{\omega} \to \mathbb{R}$, $X_n(u) = u^n$ and use Borel-measure on $[0,1]$, i.e. $P[[a,b]] = b-a$ if $0 \le a < b \le 1$. Then the c.d.f. of X_n is given by

$$\begin{aligned} F_n(x) = P[X_n \le x] &= P[\{w : X_n(\omega \le x\} \\ &= P[\{\omega : \omega^n \le x\} = P[0, x^{\frac{1}{n}}] = x^{\frac{1}{n}}. \end{aligned}$$

Thus

$$X_n(\omega] \xrightarrow[n\to\infty]{=} \begin{cases} 0 & 0 \le \omega < 1 \\ 1 & \omega = 1, \end{cases}$$

so if we define the random variable X with

$$X(\omega) = \begin{cases} 0 & 0 \le \omega < 1 \\ 1 & \omega = 1, \end{cases}$$

then obviously

$$P[|X_n - X| \geq \varepsilon] \xrightarrow[n\to\infty]{1}$$

for all $\varepsilon > 0$.
Note that we do, however, have a much stronger convergence in this example since

$$X_n(\omega) \to X(\omega) \text{ for all } \omega \in \Omega = [0,1].$$

This is *convergence of functions*, not just convergence in probability.

1.4 Almost sure convergence

1.4.1 Handout

Definition 1.4. A sequence of random variables $X_1, X_2, \ldots$ converges *almost surely* to the random variable X if

$$P\left[\lim_{n\to\infty} |X_n - X| < \varepsilon\right] = 1 \quad \forall \varepsilon > 0.$$

Note: Recall that the random variables are functions, $X_i : \Omega \to \mathbb{R}$ and we can therefore write

$$\{\omega \in \Omega : \lim_{n\to\infty} |X_n(\omega) - X(\omega)| > \varepsilon\} = A_\varepsilon.$$

We see that X_n converges almost surely to X if and only if $P[A_\varepsilon] = 0$ for all $\varepsilon > 0$.

We write $X_n \to X$ a.s.

If we define

$$A := \{\omega : X_n(\omega) \to X(\omega)\},\, A_\varepsilon := \{\omega : \lim_{n\to\infty} |X_n(\omega) - X(\omega)| < \varepsilon\}$$

then

$$A = \bigcap_{j=1}^{n} A_{1/j}$$

and we obtain

$$\begin{aligned} P[A] &= P\left[\bigcap_{j=1}^{\infty} A_{1/j}\right] \\ &= \lim_{j\to\infty} P[A_{1/j}] = 1 \end{aligned} \tag{*}$$

((*): Since A_{1_j} form a decreasing sequence of sets it is fairly easy to prove (*).) In other words, $X_n(\omega) \to X(\omega)$ except on a set $\omega \in A \subseteq \Omega$ which has probability zero. For this reason this type of convergence is commonly described as $X_n \to X$ *with probability one*.
The following has been covered:

- $X_n \xrightarrow{P} X$ if $\lim_{n\to\infty} P[|X_n - X| \geq \varepsilon] = 0$ for all $\varepsilon > 0$.

- Weak law of large numbers: $X_1, X_2, \ldots$ iid, $VX_i < \infty$ implies $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mu :=$ EX_i.

- h cont, $X_n \xrightarrow{P} X$ implies $h(X_n) \to h(X)$.

- Almost sure convergence: $X_n \to X$ a.s. if $P[\lim_{n\to\infty} |X_n - X| \geq \varepsilon] = 0$ for all $\varepsilon > 0$.

- Recall: $X_n \to X$ a.s. implies $X_n \xrightarrow{P} X$.

Theorem 1.4 (Strong law of large numbers) If $X_1, X_2, \ldots$ are i.i.d. with

$$\mathrm{E}X_i = \mu \underbrace{\mathrm{V}X_i = \sigma^2 < \infty}_{\text{not needed}}$$

and $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$, then:

$$P\left[\lim_{n\to\infty} |\bar{X}_n - \mu| < \varepsilon\right] = 1 \quad \forall \varepsilon > 0,$$

i.e. $\bar{X}_n \to \mu$ a.s. [proof omitted].

Definition 1.5. If $X_1, X_2, \ldots$ is a sequence of random variables and X is a random variable such that $F_n(x) = P[X_n \leq x]$ and $F(x) = P[X \leq x]$ satisfy $F_n(x) \to F(x)$ whenever F is continuous at x, then X_n *converges to X in distribution* , denoted $X_n \xrightarrow{D} X$.

Example 1.4. Let $X_n \sim b(n, p_n)$ where $p_n = \frac{\lambda}{n}$. We want to show that

$$X_n \xrightarrow{D} X \sim P(\lambda)$$

We have:

$$P[X_n = x] = \binom{n}{x} p_n^x (1-p_n)^{n-x} = \frac{n!}{x!\,(n-x)!\,n^x} \lambda^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n \frac{n!}{n^x\,(n-x)!}\left(1-\frac{\lambda}{n}\right)^{-x}$$

We know that $\left(1-\frac{\lambda}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-\lambda}$. We also get:

$$\frac{n!}{n^x\,(n-x)!} = \frac{n\,(n-1)\cdot\ldots\cdot(n-x+1)}{n^x} = \frac{n}{n}\cdot\frac{n-1}{n}\cdot\ldots\cdot\frac{n-x+1}{n} \xrightarrow[n\to\infty]{} 1$$

We therefore conclude that

$$P[X_n = x] = \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n \frac{n!}{n^x\,(n-x)!}\left(1-\frac{\lambda}{n}\right)^{-x} \xrightarrow[n\to\infty]{} e^{-\lambda}\frac{\lambda^x}{x!} = P[X = x],$$

where $X \sim P(\lambda)$.
Since we have shown that $lim_{n\to\infty} P[X_n = x] = P[X = x]$, we also see that $lim_{n\to\infty} P[X_n \leq x] = P[X \leq x]$ (these are finite sums and each element converges).
It follows that the sequence X_n converges in distribution to X, or

$$X_n \xrightarrow{D} X \sim P(\lambda).$$

Theorem 1.5 $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$ [see exercise 5.40].

Theorem 1.6 $X_n \xrightarrow{D} c \Rightarrow X_n \xrightarrow{P} c$ if $c \in \mathbb{R}$.

2 Order statistics

2.1 Order statistics

2.1.1 Handout

Suppose $X_1, \dots, X_n$ are i.i.d., i.e. are a random sample.

Definition 2.1. Define the random variable $X_{(n)} := \max\{X_1, \dots, X_n\}$.

Note 2.1. Sometimes (n) is defined as the random variable which corresponds to the largest element in $(X_1, \dots, X_n)$.

Definition 2.2. We define $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ to be be the n *order statistics* of the random sample $X_1, \dots, X_n$.

Note: Formally, since each random variable is really a function, these new variables need to be defined as new functions...

Example 2.1. If $X_i \sim U(0,1)$ then we have for $0 \leq \omega \leq 1$:

$$\begin{aligned} P[X_{(n)} \leq \omega] &= P[X_1 \leq \omega_1, \dots, X_n \leq \omega] \\ &= P[X_1 \leq \omega]^n \quad (iid) \\ &= \omega^n \xrightarrow[n\to\infty]{} \begin{cases} 0 & 0 \leq \omega < 1 \\ 1 & \omega = 1 \end{cases} \end{aligned}$$

so that $X_{(n)} \xrightarrow{D} X$ with $P[X = 1] = 1$, i.e. $X_{(n)} \xrightarrow{D} 1$, and it follows that

$$P[X \leq x] = \begin{cases} 0 & x < 1 \\ 1 & x \geq 1 \end{cases}.$$

Note:

$$\begin{aligned} P[X_{(1)} \leq \omega] &= 1 - P[X_{(1)} > \omega] = 1 - P[X_1 > \omega]^n \\ &= 1 - (1-w)^n \xrightarrow[n\to\infty]{} \begin{cases} 0 & \omega = 0 \\ 1 & 0 < \omega \leq 1 \end{cases} \end{aligned}$$

so that $X_{(1)} \xrightarrow{D} 0$.

We also obtain:

$$
\begin{aligned}
P[|X_{(n)}-1| \leq \varepsilon] &= P[1-\varepsilon \leq X_{(n)} \leq 1+\varepsilon] \\
&= P[X_{(n)} \geq 1-\varepsilon] = 1 - P[X_{(n)} \leq 1-\varepsilon] \\
&= 1-(1-\varepsilon)^n \xrightarrow[n\to\infty]{1}
\end{aligned}
$$

if $0 < \varepsilon < 1$, and hence $X_{(n)} \xrightarrow{P} X$. We have $X'_{(n)} \xrightarrow{D} 1$ and $X_{(n)} \xrightarrow{P} 1$.
The density of $X_{(n)}$ is given by

$$
\begin{aligned}
f_n(x) = F_n'(x) &= \frac{d}{dx}F(x)^n \\
&= nf(x)F(x)^{n-1} = nx^{n-1}I_{[0,1]}(x).
\end{aligned}
$$

The expected value of $X_{(n)}$ is therefore

$$
\mathrm{E}X_{(n)} = \int_0^1 xnx^{n-1}\,dx = \ldots = \frac{n}{n+1} \xrightarrow[n\to\infty]{} 1,
$$

and the variance is obtained by first evaluating

$$
E[X_{(n)}^2] = \int_0^1 x^2 nx^{n-1}\,dx = \ldots = \frac{n}{n+2}
$$

from which we see that

$$
\mathrm{V}[X_{(n)}] = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 = \frac{n}{(n+1)^2(n+2)},
$$

i.e. $\mathrm{V}[X_{(n)}]$ “behaves like” $\frac{1}{n^2}$.
Since $X_{(n)}$ converges to 1 in distribution and the standard deviation behaves like $1/n$, i is of interest to see what happens to the distribution of the random variable $\frac{X_{(n)}-1}{1/n}$ or simply $n(1-X_{(n)})$. We would expect this transformed random variable to have (approximately) mean zero and variance one, so it should converge to a proper non-constant random variable.
We obtain:

$$
P[n(1-X_{(n)}) \leq t] \xrightarrow[n\to\infty]{} 1-e^{-t}
$$

(this is a popular exam question).

3 Random number generation

3.1 Continuous distributions

3.1.1 Handout

Let $U \sim U(0,1)$. If F is increasing, continuous and

$$0 \leq F(x) \leq 1, x \in \mathbb{R}.$$
$$F(x) \xrightarrow[x\to\infty]{} 0,$$
$$F(x) \xrightarrow[x\to 0]{} 1,$$

and we set

$$Y := F^{-1}(U)$$

then we see that

$$P[Y \leq y] = P[F^{-1}(U) \leq y] = P[U \leq F(y)] = F(y),$$

so that $Y \sim F$.

Example 3.1 (Example of usage). If $U \sim U(0,1)$ and

$$\underbrace{\Phi(x)}_{\text{pnorm(x)}inR} := \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-t^2/2}}_{\text{dnorm(t)}} dt,$$

then

$$\Phi^{-1}(U) \sim \underbrace{n(0,1)}_{\text{rnorm(1) in R}} .$$

Note: Recall that we can write

$$g(x) = \sum_{i=0}^{\infty} \frac{g^{(i)}(a)}{i!}(x-a)^i, \quad x \in (a-r, a+r)$$

if g if infinitely differentiable and $g^{(n)}(x)$ disappears "fast enough" as $n \to \infty$ [specifically $\exists A > 0$ s.t. $g^{(n)}(x) \leq A^n \forall n$].

3.2 Discrete distributions

3.2.1 Handout

Discrete distributions:
Define $F^{-1}(u) := inf\{x : F(x) \geq u\}$ and note that if F is a c.d.f. then F is continuous from the right so the infimum is a minimum.
Suppose F "jumps" at x, so that $P[X = x] > 0$, i.e. $F(x_-) < F(x_+) = F(x)$, then $F(x) < u \leq F(x) \Rightarrow F^{-1}(u) = x$. In that case $X := F^{-1}(U)$ has a point mass probability of $P[X = x]$ at x.

4 Central limit theorem

4.1 Lemma on m.g.f.s and c.d.f.s

4.1.1 Handout

Lemma
If X_n each have c.d.f. F_n and m.g.f. M_n, defined in $]-h,h[$ and there is a c.d.f. F which corresponds to m.g.f. M and $M_n(t) \underset{n\to\infty}{\to} M(t)$ for $|t| < h$ then $X_n \xrightarrow{D} X$ if X has c.d.f. F.
Note: A corresponding lemma holds for characteristic functions.

4.2 A note on Taylor series

4.2.1 Handout

Recall that we can write

$$g(x) = \sum_{i=0}^{\infty} \frac{g^{(i)}(a)}{i!}(x-a)^i,\ x \in]a-r, a+r[$$

if g is infinitely differentiable and $g^{(n)}(x)$ disappears "fast enough"as $n \to \infty$ (i.e. $\exists A > 0$ s.t. $g^{(n)}(x) \leq A^n$).

4.3 A lemma on limits

4.3.1 Handout

If (a_n) is a sequence of numbers s.t. $a_n \to 0$ then $\lim_{n\to\infty} \left(1 + \frac{x+a_n}{n}\right)^n = e^x$

4.4 Central limit theorem

4.4.1 Handout

Theorem 4.1 (Central limit theorem, CLT) Let $X_1, X_2, \ldots$ be iid random variables such that the common moment generating function M exists in a neighborhood of 0. Let $\mathrm{E}X_i = \mu$, $\mathrm{V}X_i = \sigma^2 > 0$ and define $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. If

$$G_n(x) := P\left[\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right]$$

then

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt,$$

i.e. if $Z \sim n(0,1)$ then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} Z.$$

Proof. Assume that $M(t) = E[e^{tX}]$ exists for $|t| < h$. Define $Y_i = \frac{X_i - \mu}{\sigma}$ and let Y be a random variable with the same distribution as all Y, so the m.g.f. of Y is

$$\begin{aligned} M_Y(t) =& E[e^{tY}] = E[e^{tY_i}] = E[e^{t\frac{X_1-\mu}{\sigma}}] \\ =& E[e^{\frac{t}{\sigma}X_1} e^{-\frac{\mu}{\sigma}t}] = e^{-t\frac{\mu}{\sigma}} E[e^{\frac{t}{\sigma}X_1}] = e^{-t\frac{\mu}{\sigma}} M\left(\frac{t}{\sigma}\right) \end{aligned}$$

which exists for $|t| < h\sigma$.
Now define

$$\begin{aligned} Z_n := \frac{X_n - \mu}{\sigma/\sqrt{n}} =& \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)}{\sigma/\sqrt{(n)}} \\ =& \frac{\sqrt{(n)}}{n}\sum_{i=1}^{n} = \frac{1}{\sqrt{(n)}}\sum_{i=1}^{n} Y_i \end{aligned}$$

Next look at the m.g.f of Z_n

$$\begin{aligned} M_{Z_n}(t) =& E\left[e^{\frac{t}{\sqrt{(n)}}}\sum_{i=1}^{n} Y_i\right] \\ =& E\left[e^{\frac{t}{\sqrt{(n)}}Y_1} e^{\frac{t}{\sqrt{(n)}}Y_2} \ldots e^{\frac{t}{\sqrt{(n)}}Y_n}\right] \\ =& \prod_{i=1}^{n} E\left[e^{\frac{t}{\sqrt{(n)}}}Y_i\right] \\ =& \left(E\left[e^{\frac{t}{\sqrt{(n)}}Y_1}\right]\right)^n \\ =& M_Y\left(\frac{t}{\sqrt{(n)}}\right)^n \end{aligned}$$

which exists if $\left|\frac{t}{\sqrt{(n)}}\right| < h\sigma$.
Now we use the note on Taylor series to write

$$M_Y\left(\frac{t}{\sqrt{(n)}}\right) = \sum_{k=1}^{\infty} M_Y^k(0)\frac{(t/\sqrt{(n)})^k}{k!}$$

which holds if $|t| < h\sigma\sqrt{(n)}$. Recall that $M_Y(0) = 1$, $M_T'(0) = E[Y] = 0$, $M_Y''(0) = E[Y^2] = 1$ and we can write the series as the first parts plus a remainder such as

$$M_Y(\frac{t}{\sqrt{(n)}} = 1 + 0 + 1\frac{\left(t/\sqrt{(n)}\right)^2}{t!} + R\left(\frac{t}{\sqrt{(n)}}\right)$$

where R is the remainder that satisfies

$$\frac{R(x)}{x^2} \underset{x\to 0}{\to} 0 \text{ i.e. } \frac{t}{\left(t/\sqrt{(n)}\right)^2} \underset{n\to\infty}{\to} 0$$

[Note: We do not use the full Taylor expansion].

Next consider the limit of m.g.fs

$$\lim_{n\to\infty} M_Y\left(\frac{t}{\sqrt{n}}\right)^n = \lim_{n\to\infty}\left[1+\frac{t^2}{2n}+2\frac{2}{\sqrt{n}}\right]^n$$
$$= \lim_{n\to\infty}\left[1+\frac{t^2/2+2n(t/\sqrt{n})}{n}\right]^n$$
$$= \lim_{n\to\infty}\left[1+\frac{t^2/2+a_n}{n}\right]^2$$

where a_n is a sequence which satisfies $a_n \underset{n\to\infty 0}{\to} 0$. According to lemma we obtain

$$\lim_{n\to\infty} M_Z(t) = e^{t^2/2}$$

and this holds for $t \in \mathcal{R}$.
If $Z \sim n(0,1)$ then $M_Z(t) = e^{t^2/2}$, i.e. $M_{Z_n}(t) \to M_Z(t)$ and therefore $Z_n \xrightarrow{D} Z$ i.e.

$$\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z \sim n(0,1).$$

□

We have looked at

- Almost sure convergence
- Convergence in probability
- Convergence in distribution

This is always based on a sequence $X_1, X_2, \ldots$ (not always independent) e.g.

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$Y_n \xrightarrow{a.s.} \mu = \mathrm{E}[X_i]$$

if

$$\mathrm{V}[X_i] < \infty$$

such that

$$Y_n \xrightarrow{\mathcal{P}} \mu$$

We now have

$$\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim n(0,1)$$
$$X_1, X_2, \ldots \text{ iid}$$
$$V[X_i] < \infty$$

This last conclusion was obtained by looking at the moment generating function of Z_n, where

$$Z_n = \sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}.$$

$$
\begin{aligned}
M(t) &= \mathrm{E}\left[e^{tX}\right], && \text{og með Taylor-liðun:}\\
&= \mathrm{E}\left[1 + \frac{tX}{1!} + \frac{tX^2}{2!} + \cdots\right],\\
M_{Z_n}(t) &= \mathrm{E}\left[\exp\left[t\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}\right]\right]\\
&= \mathrm{E}\left[\exp\left[t\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i - \mu}{\sigma}\right]\right]\\
&= \mathrm{E}\left[\prod_{i=1}^{n}\exp\left[t\frac{1}{\sqrt{n}}\frac{X_i - \mu}{\sigma}\right]\right]\\
&= \left(\mathrm{E}\left[\exp\left[t\frac{1}{\sqrt{n}}\frac{X - \mu}{\sigma}\right]\right]\right)^n && \text{(iid)}\\
&= \left(\mathrm{E}\left[1 + \frac{1}{1!}\left(\frac{t}{\sqrt{n}}\frac{X - \mu}{\sigma}\right) + \frac{1}{2!}\left(\frac{t}{\sqrt{n}}\frac{X - \mu}{\sigma}\right)^2 + \cdots\right]\right)^n\\
&\approx \left(\mathrm{E}\left[1 + \frac{1}{2!}\left(\frac{t}{\sqrt{n}}\right)^2\left(\frac{X - \mu}{\sigma}\right)^2\right]\right)^n\\
&= \left(1 + \frac{t^2}{2n}\cdot 1\right)^n \xrightarrow[n\to\infty]{} e^{\frac{t^2}{2}}.
\end{aligned}
$$

4.5 Slutsky's theorem

4.5.1 Handout

Theorem 4.2 (Slutzky) If

$$X_n \xrightarrow{\mathcal{D}} X \text{ og } Y_n \xrightarrow{\mathcal{P}} a,$$

then

$$X_n Y_n \xrightarrow{\mathcal{D}} aX \text{ og } X_n + Y_n \xrightarrow{\mathcal{D}} a + X.$$

Example 4.1. We know that if $X_n \sim b(n, p)$ then

$$\hat{p}_n := \frac{X_n}{n} \xrightarrow{\mathcal{D}} p$$

and we know that the function

$$x \mapsto \sqrt{x(1-x)}$$

is continuous so that

$$\sqrt{\hat{p}_n(1-\hat{p}_n)} \xrightarrow{\mathcal{P}} \sqrt{p(1-p)}$$

We also know that X_n can be written as a sum

$$X_n \overset{\mathcal{D}}{=} \sum_{i=1}^{n} Y_i$$

where Y_i are independent and Bernoulli, $Y_i \sim b(1,p)$ i.i.d. and $\hat{p}_n$ therefore has the same distribution as an average,

$$\hat{p}_n \stackrel{\mathcal{D}}{=} \frac{\sum_{i=1}^{n} Y_i}{n},$$

so

$$\frac{\hat{p}_n - E[\hat{p}]}{\sqrt{V[\hat{p}]}} \xrightarrow{\mathcal{D}} n(0,1)$$

But $V[\hat{p}] = \frac{p(1-p)}{n}$ and so we can use Slutsky's theorem to conclude

$$\frac{\hat{p}_n - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \xrightarrow{\mathcal{D}} n(0,1)$$

On assumptions:
1) When should we use t-distribution?

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$$

This holds exactly if $X_1,..,X_n \sim n\left(\mu,\sigma^2\right)$, iid.
2) But if n is "large"then this still holds as an approximation, based on combining the CLT and Slutzky's theorem:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim\dot{\sim} n(0,1)$$

Here we just need X_i iid with finite σ^2 – we do **not** need the original random variables to be Gaussian.
Slutzky's theorem has a series of consequences. If $X_1, X_2, \ldots$ are iid with

$$\mathrm{E}\left[X^2\right] < \infty$$

(so that $\sigma^2 = V[X] < \infty$) then the mean $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ has the property that

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} n(0,1)$$

and we also know that

$$\mathrm{S}_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

Further, $\mathrm{S}_n^2 \xrightarrow{\mathcal{P}} \sigma^2$ and hence $S_n \xrightarrow{\mathcal{P}} \sigma$ so Slutzky's theorem implies:

$$\begin{aligned}\frac{\bar{X}_n - \mu}{\mathrm{S}_n/\sqrt{n}} &= \frac{\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}}{\mathrm{S}_n/\sigma} \\ &= \underbrace{\frac{\sigma}{\mathrm{S}_n}}_{\xrightarrow{\mathcal{P}} 1}\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{D}} n(0,1).\end{aligned}$$

Note that this implies that we can approximate probabilities of events such that

$$\mathrm{P}\left[\bar{X}_n - \kappa\frac{\mathrm{S}_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \kappa\frac{\mathrm{S}_n}{\sqrt{n}}\right] = \mathrm{P}\left[-\kappa \leq \frac{\bar{X}_n - \mu}{\mathrm{S_n}/\sqrt{n}} \leq \kappa\right]$$

by corresponding $n(0,1)$ probabilities, i.e.

$$\mathrm{P}\left[\bar{X}_n - \kappa\frac{\mathrm{S}_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \kappa\frac{\mathrm{S}}{\sqrt{n}}\right] \approx 1 - \alpha$$

where $\kappa = z_{1-\frac{\alpha}{2}}$. This is an **approximation**.
Finally, if $X_i \sim n(\mu, \sigma^2)$ iid already know that

$$T_n := \frac{\bar{X}_n - \mu}{\mathrm{S}/\sqrt{n}} = \frac{\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \Big/ (n-1)}}$$

and $\dfrac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$ so that

$$P\left[\bar{X}_n - \kappa\frac{\mathrm{S}}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \kappa\frac{\mathrm{S}}{\sqrt{n}}\right] = 1 - \alpha$$

where $\kappa = t_{n-1,1-\frac{\alpha}{2}}$. This is **exact** but requires the assumption of normality of the data.

Example 4.2. $X_i = \begin{cases} 0 \\ 1 \end{cases}, P[X_i = 1] = p = 1 - P[X_i = 0], X_i$ iid, i.e. $X_i \sim b(1,p)$ iid and $Y_n := \sum_{i=1}^n X_i \sim b(n,p)$.

We know that $\frac{\frac{1}{n}Y_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} n(0,1)$ (CLT) since $\mu = \mathrm{E}[Y_n]/n = p$ and $\sigma = \mathrm{V}\left[\frac{Y_n}{n}\right] = \frac{1}{n^2}np(1-p)$ i.e. if $\hat{p}_n = \frac{1}{n}Y_n$ then

$$\frac{\hat{p} - p}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{D}} n(0,1)$$

We could use $\mathrm{P}\left[-z_{1-\frac{\alpha}{2}} \leq \frac{\hat{p}-p}{\sqrt{np(1-p)}} \leq z_{1-\frac{alpha}{2}}\right] \approx 1 - \alpha$ to obtain intervals of the form

$$\mathrm{P}\left[f_1(\hat{p}) \leq p \leq f_2(\hat{p})\right] \approx 1 - \alpha,$$

but since we know that $\hat{p}_n \xrightarrow{\mathcal{P}} p$ we obtain using Slutzky's theorem

$$\frac{\hat{p} - p}{\sqrt{n\hat{p}(1-\hat{p})}} \xrightarrow{\mathcal{D}} n(0,1) \tag{1}$$

[more exactly: $\hat{p} \xrightarrow{\mathcal{P}} p$ and $s \mapsto \frac{1}{\sqrt{s(1-s)}}$ is continuous

$$\Rightarrow \frac{1}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{\mathcal{P}} \frac{1}{\sqrt{p(1-p)}}$$

and (1) is therefore a consequence of Slutzky's theorem]
i.e. we obtain:

$$\mathrm{P}\left[\hat{p} - z_{1-\frac{\alpha}{2}}\sqrt{n\hat{p}(1-\hat{p})} \leq p \leq \hat{p} + z_{1-\frac{\alpha}{2}}\sqrt{n\hat{p}(1-\hat{p})}\right] \approx 1 - \frac{\alpha}{2}$$

4.6 The Delta method

4.6.1 Handout

Theorem 4.3 (Delta method, (5.5.24)) *Let $Y_1, Y_2 \ldots$ be a sequence of random variables such that*

$$\sqrt{n}(Y_n - \theta) \xrightarrow{\mathcal{D}} n(0, \sigma^2)$$

and assume that g is a function such that $g'(\theta) \neq 0$. Then:

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{\mathcal{D}} n\left(0, (g'(\theta))^2 \sigma^2\right).$$

Note: $g(Y_n) = g(\theta) + g'(\theta)\frac{Y_n-\theta}{1!} + g''(\theta)\frac{(Y_n-\theta)^2}{2!} + \cdots$ so we can "approximate" $\mathrm{V}[g(Y_n)]$ með $\mathrm{V}[g(Y_n)] = \mathrm{E}[(g(Y_n) - g(\theta))^2] \approx \mathrm{E}\left[(g'(\theta)(Y_n - \theta))^2\right]$

Example 4.3. Recall that

$$\sqrt{n}(\hat{p} - p) \xrightarrow{\mathcal{D}} n(0, p(1-p))$$

since

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$\mathrm{V}\left[\sqrt{n}\hat{p}\right] = n\frac{p(1-p)}{n} = p(1-p).$$

This can now be used to derive the properties of the **arc sine square root transformation**

$$\arcsin\sqrt{\hat{p}} \xrightarrow{\mathcal{D}} ?$$

Example 4.4 (5.5.25). Assume that $(\bar{X}_n - \mu)\sqrt{n} \xrightarrow{\mathcal{D}} n(0, \sigma^2)$ and $\mu \neq 0$. Consider the function $g(\mu) := \frac{1}{\mu}$ with $g'(\mu) = \frac{1}{\mu^2}$ to obtain

$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right) \xrightarrow{\mathcal{D}} n\left(0, \frac{\sigma^2}{\mu^4}\right)$$

but of course we would prefer a random variable which is not a function of σ^2, e.g.:

$$\frac{\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right)}{S_n/\bar{X}_n^2} \xrightarrow{\mathcal{D}} n(0,1)$$

and we obtain by applying a few theorems:

$$\begin{cases} \bar{X}_n \xrightarrow{\mathcal{P}} \mu \\ S_n \xrightarrow{\mathcal{P}} \sigma \end{cases} \Rightarrow \begin{cases} \bar{X}_n^2 \xrightarrow{\mathcal{P}} \mu^2 \\ \frac{1}{S_n} \xrightarrow{\mathcal{P}} \frac{1}{\sigma} \end{cases} \Rightarrow \frac{\bar{X}_n^2}{S_n} \xrightarrow{\mathcal{P}} \frac{\mu^2}{\sigma}.$$

This included using Slutzki with

$$\frac{\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right)}{\sigma/\mu^2} \xrightarrow{\mathcal{D}} n(0,1).$$

Sönnun. Recall Slutsky's theorem: If $X_n \to X$ in distribution and $Z_n \to a$, a constant, then: $X_n Z_n \to aX$ in distribution, and $X_n + Y_n \to X_n + a$ in distribution

Now, the Taylor expansion of $g(Y_n)$ around $Y_n = \theta$ is

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + R$$

where R is the remainder and $R \to 0$ as $Y \to \theta$. From the assumption that Y_n satisfies the standard Central Limit Theorem, we have $Y_n \to \theta$ in probability, so it follows that $R \to 0$ in probability as well. Rearranging the terms we have:

$$\sqrt{n}(g(Y_n) - g(\theta)) = g'(\theta)\sqrt{n}(Y_n - \theta) + R$$

Applying Slutsky's theorem with X_n as $g'(\theta)\sqrt{n}(Y_n - \theta)$ and Z_n as R, we have the right hand side converging to $n\left(0, \sigma^2 g'(\theta)^2\right)$.