Random Sampling*

Ricardo Dahis

PUC-Rio, Department of Economics

Summer 2023

*Last updated: January 19, 2023, 19:13h

Contents

Contents

definition

- definition: $(X_1, \ldots, X_n)$ is a random sample of size n from the population $f_X(x)$ if they are mutually independent random variables with the same marginal pmf/pdf given by $f_X(x)$.

- alternatively, we say that $X_1, \ldots, X_n$ are independent and identically distributed (iid) with pmf/pdf $f_X(x)$

$$f_{\boldsymbol{X}}(x_1, \ldots, x_n|\boldsymbol{\theta}) \;=\; f_X(x_1|\boldsymbol{\theta}) \cdots f_X(x_n|\boldsymbol{\theta}) \;=\; \prod_{i=1}^{n} f_X(x_i|\boldsymbol{\theta})$$

- statistical setting: we assume that the population we observe belongs to a given parametric family, but the true parameter value is unknown.

joint pdf of an exponential sample

- let $X_1, \dots, X_n$ form a random sample from an exponential distribution with parameter λ, then the joint pdf reads

$$f_{\mathbf{X}}(x_1, \dots, x_n|\lambda) = \prod_{i=1}^{n} f_X(x_i|\lambda) = \prod_{i=1}^{n} \frac{1}{\lambda} e^{-x_i/\lambda} = \frac{e^{-\sum_{i=1}^{n} x_i/\lambda}}{\lambda^n}$$

- example: what is the probability of all X_i last more than 2 years?

$$\begin{aligned} \mathbb{P}(X_1 > 2, \dots, X_n > 2|\lambda) &= \mathbb{P}(X_1 > 2|\lambda) \cdots \mathbb{P}(X_n > 2|\lambda) \\ &= [\mathbb{P}(X_1 > 2|\lambda)]^n \\ &= \left(e^{-2/\lambda}\right)^n = e^{-2n/\lambda} \end{aligned}$$

sampling from an infinite population

- independence assumption implies that drawing X_i does not affect the distribution of X_j and hence the latter is from the same population
  - it is as if the population were infinite
- finite populations: data collection now matters in that the iid assumption may not hold depending on how one samples from the population is with vs without replacement
- examples:
  - (i) bootstrap employs a resampling scheme with replacement
  - (ii) no replacement kills independence, $\mathbb{P}(X_i = x | X_j = x) = 0$ but with independence $\mathbb{P}(X_i = x) = \mathbb{P}(X_j = x)$

near independence

- definition: $X_1, \ldots, X_n$ are nearly independent if population size is large enough and hence one may evoke random sampling as an approximation

- example: $\mathbb{P}(X_i = x | X_j = x_j) = \frac{1}{n-1} \cong \mathbb{P}(X_i = x | X_j = x) = 0$ for n large enough

- example: draw a sample $\{X_1, \ldots, X_{10}\}$ without replacement from a discrete uniform population $\{1, \ldots, 1000\}$ (hypergeometric distribution)

$$\begin{aligned}
\mathbb{P}(X_1 > 200, \ldots, X_{10} > 200) &= \frac{\binom{800}{10}\binom{200}{0}}{\binom{1000}{10}} = 0.106164 \\
&\cong \mathbb{P}(X_1 > 200) \cdots \mathbb{P}(X_{10} > 200) \\
&= 0.8^{10} = 0.107374
\end{aligned}$$

Contents

statistic

- we usually compute some value after a sample $X_1, \ldots, X_n$ is drawn.
- definition: let $(X_1, \ldots, X_n)$ denote a random sample of size n from a population, then the random vector $Y = T(X_1, \ldots, X_n)$ is a statistic if it is a vector-valued function of $X_1, \ldots, X_n$ whose domain includes the sample space of $X_1, \ldots, X_n$
  - the definition is very broad, but restriction is that Y cannot be a function of parameters.
- because random samples have a simple probabilistic structure, the sampling distribution of $T(X_1 \ldots, X_n)$ is particularly tractable.

statistic

- examples:
  - $T(x_1, \ldots, x_n) = 1$
  - $T(x_1, \ldots, x_n) = x_1$
  - $T(x_1, \ldots, x_n) = \max\{x_1, \ldots, x_n\}$ (maximum)
  - $T(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}_n$ (sample mean)
  - $T(x_1, \ldots, x_n) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = s_n^2$ (sample variance)
  - $T(x_1, \ldots, x_n) = \sqrt{s_n^2} = s_n$ (sample standard deviation)
- note that we often write $T = T(x_1, \ldots, x_n)$
- functions of random variables are themselves random variables: we write $\bar{X}_n$ and $\bar{x}_n$ for a particular realized value.

sample mean, variance, and standard deviation

- theorem (CB 5.2.4): let $x_1, \ldots, x_n$ denote any real numbers and let $\bar{x}_n \equiv \frac{1}{n}\sum_{i=1}^n x_i$, then

  (i) $\min_a \sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2$

  (ii) $(n-1)s_n^2 \equiv \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \sum_{i=1}^n x_i^2 - n\,\bar{x}_n^2$

- proof of (i):

$$
\begin{aligned}
\sum_{i=1}^n (x_i - a)^2 &= \sum_{i=1}^n (x_i - \bar{x}_n + \bar{x}_n - a)^2 \\
&= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + \underbrace{2\sum_{i=1}^n (x_i - \bar{x}_n)(\bar{x}_n - a)}_{(\bar{x}_n - a)\sum_{i=1}^n (x_i - \bar{x}_n) = 0} + \sum_{i=1}^n (\bar{x}_n - a)^2 \\
&= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + \sum_{i=1}^n (\bar{x}_n - a)^2
\end{aligned}
$$

which is minimized when $a = \bar{x}$.

- proof of (ii): taking $a = 0$,

$$\begin{aligned}\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + \sum_{i=1}^{n} \bar{x}_n^2 \\ &= \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + n\bar{x}_n^2\end{aligned}$$

$\blacksquare$

sample mean, variance, and standard deviation

- theorem (CB 5.2.5): let $X_1, \ldots, X_n$ form a random sample from a population and let $g(x)$ be a function such that $\mathbb{E}[g(X)]$ and $\mathrm{var}[g(x)]$ exist, then

  (i) $\mathbb{E}\left[\sum_{i=1}^n g(X_i)\right] = n\,\mathbb{E}[g(X_1)]$

  (ii) $\mathrm{var}\left[\sum_{i=1}^n g(X_i)\right] = n\,\mathrm{var}[g(X_1)]$

- proof: note that

$$\mathbb{E}\left(\sum_{i=1}^n g(X_i)\right) = \sum_{i=1}^n \mathbb{E}g(X_i) \overset{iid}{=} \sum_{i=1}^n \mathbb{E}g(X_1) = n \cdot \mathbb{E}g(X_1)$$

for the second part,

$$\begin{aligned}\mathrm{var}\left(\sum_{i=1}^n g(X_i)\right) &= \mathbb{E}\left[\sum_{i=1}^n g(X_i) - \mathbb{E}\left(\sum_{i=1}^n g(X_i)\right)\right]^2 \\ &= \mathbb{E}\left[\sum_{i=1}^n g(X_i) - \sum_{i=1}^n \mathbb{E}g(X_i)\right]^2\end{aligned}$$

sample mean, variance, and standard deviation

- proof (cont'd):

$$\mathbb{E}\left[\sum_{i=1}^n g(X_i) - \sum_{i=1}^n \mathbb{E}g(X_i)\right]^2 = \mathbb{E}\left[\sum_{i=1}^n g(X_i) - \mathbb{E}g(X_i)\right]^2 = \mathbb{E}\left[\sum_{i=1}^n h_i\right]^2$$

denoting $h_i \equiv g(X_i) - \mathbb{E}g(X_i)$. Then

$$\mathbb{E}\left[\sum_{i=1}^n h_i\right]^2 = \sum_{i=1}^n \mathbb{E}h_i^2 + \sum_{i=1}^n \sum_{j=1, j\neq i}^n \mathbb{E}(h_i h_j)$$

but $\mathbb{E}(h_i h_j) = \mathbb{E}\left([g(X_i) - \mathbb{E}g(X_i)][g(X_j) - \mathbb{E}g(X_j)]\right) = \mathrm{cov}(g(X_i), g(X_j)) = 0$. It follows that

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^n h_i\right]^2 &= \sum_{i=1}^n \mathbb{E}h_i^2 = \sum_{i=1}^n \mathbb{E}(g(X_i) - \mathbb{E}g(X_i))^2 \\ &= \sum_{i=1}^n \mathrm{var}(g(X_i)) \overset{iid}{=} \sum_{i=1}^n \mathrm{var}(g(X_1)) = n \cdot \mathrm{var}(g(X_1))\end{aligned}$$

$\blacksquare$

sample mean, variance, and standard deviation

- theorem (CB 5.2.6): if the population has mean μ and variance σ^2, then

  (i) $\mathbb{E}(\bar{X}_n) = \mu$ — unbiasedness

  (ii) $\text{var}(\bar{X}_n) = \sigma^2/n$ — precision

  (iii) $\mathbb{E}(S_n^2) = \sigma^2$ — unbiasedness

- proof (i):

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{n}{n}\mathbb{E}(X_1) = \mu$$

- proof (ii):

$$\text{var}(\bar{X}_n) = \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\text{var}\left(\sum_{i=1}^{n} X_i\right) = \frac{n}{n^2}\text{var}(X_1) = \frac{\sigma^2}{n}$$

- proof (iii):

$$
\begin{aligned}
\mathbb{E}(S^2) &= \mathbb{E}\left(\frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]\right) \\
&\overset{iid}{=} \frac{1}{n-1}\left(n\mathbb{E}X_1^2 - n\mathbb{E}\bar{X}^2\right) \\
&= \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) \\
&= \frac{1}{n-1}\left(n\sigma^2 - \sigma^2\right) \\
&= \sigma^2
\end{aligned}
$$

which completes the proof. ■

- definition: we say that the statistic T is *unbiased* for the parameter θ if $\mathbb{E}(T) = \theta$.
- according to the example above, $\bar{X}_n$ is unbiased for μ and S_n^2 is unbiased for σ^2.
- we will now discuss in more detail the distribution of $\bar{X}_n$.

sampling distribution of the mean

- theorem (CB 5.2.7): let $(X_1, \ldots, X_n)$ be a random sample from a population with pdf $f_X(x)$ and mgf $M_X(t)$ and denote $Y = X_1 + \cdots + X_n$. Then

$$\begin{aligned} f_{\bar{X}_n}(x) &= n f_Y(nx) \\ M_{\bar{X}_n}(t) &= [M_X(t/n)]^n \end{aligned}$$

- proof: the first result is rather mechanical since $\bar{X}_n = n^{-1}Y$ and applying the change-of-variable theorem. For the latter, apply the theorem that if $X_1, \ldots, X_n$ are independent, then for $Z = \sum_{i=1}^n a_i X_i + b_i$,

$$M_Z(t) = \left(e^{t \sum b_i}\right) \prod_{i=1}^n M_{X_i}(a_i t)$$

so

$$M_{\bar{X}}(t) = \prod_{i=1}^n M_{X_i}\left(\frac{1}{n} t\right) \overset{iid}{=} \left[M_X\left(\frac{1}{n} t\right)\right]^n$$

■

- example: let $X_1, \ldots, X_n$ form a random sample from a normal distribution with mean μ and variance σ^2, then the mgf of the sample mean is

$$\begin{aligned} M_{\bar{X}_n}(t) &= [M_X(t/n)]^n = \left[\exp\left(\frac{\mu t}{n} + \frac{\sigma^2 (t/n)^2}{2}\right)\right]^n \\ &= \exp\left(\mu t + \frac{(\sigma^2/n)\, t^2}{2}\right) \end{aligned}$$

and hence $\bar{X}_n \sim N(\mu, \sigma^2/n)$

sampling from a location-scale family

- let $(X_1, \ldots, X_n)$ denote a random sample from a location-scale family $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, then the distribution of $\bar{X}_n$ has a simple relationship with the distribution of the sample mean $\bar{Z}_n$ of a random sample from the standard family distribution $f(z)$

- how?

  (i) there exist random variables $Z_1, \ldots, Z_n$ such that $X_i = \sigma Z_i + \mu$

  (ii) $Z_1, \ldots, Z_n$ are also mutually independent

  (iii) $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_{i=1}^{n} (\sigma Z_i + \mu) = \sigma \bar{Z}_n + \mu$

  (iv) if $\bar{Z}_n \sim g(z)$, then $\bar{X}_n \sim \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right)$

- example: if $(X_1, \ldots, X_n)$ is a random sample from a Cauchy(μ, σ^2), then $\bar{X}_n \sim$ Cauchy(μ, σ^2) as well

Contents

sample mean and variance

- theorem: let $(X_1, \ldots, X_n)$ be a random sample from a $N(\mu, \sigma^2)$ population, then
  - (i) $\bar{X}_n \sim N(\mu, \sigma^2/n)$
  - (ii) $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1}$
  - (iii) $\bar{X}_n$ and S_n^2 are independent random variables
  - (iv) $\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}$
- proof (i): already established

sample mean and variance

- before moving ahead with the proof of (ii), let's establish some facts about quadratic forms
- definition: let Z be a n-dimensional vector of independent random normal variables. Then

$$Z'Z = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

  and the pdf of a χ_p^2 is $f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}$

- let $X \sim N(0, \Sigma)$. Then

$$X'\Sigma^{-1}X = X'\Sigma^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}X = Z'Z \sim \chi_n^2$$

  since $\Sigma^{-\frac{1}{2}}X \sim N(0, I)$.

- theorem: let P be an m-dimensional orthogonal projection matrix in $\mathbb{R}^n$. That is, $P^2 = P$ (projection matrix) and $P'P = PP' = I$ and $P' = P^{-1}$ (orthogonal matrix) then $Z'PZ \sim \chi_m^2$ with $Z \sim N(0, I)$.

- proof (ii): define $P_\iota = \iota(\iota'\iota)^{-1}\iota' = \frac{\iota\iota'}{n}$, where ι is the n-dimensional vector of ones. Let $M = I - P_\iota$ be the annihilator matrix. Note that
  - P_ι is symmetric (verify)
  - P_ι is a projection matrix: $P_\iota^2 = \frac{\iota\iota'}{n}\frac{\iota\iota'}{n} = \frac{\iota\iota'\iota\iota'}{n^2} = \frac{\iota\iota'}{n} = P_\iota$
  - $MX = (I - P_\iota)X = X - \iota\bar{X}$
  - $M'M = (I - P_\iota)'(I - P_\iota) = (I - P_\iota) = M$
  - $P_\iota X = \iota\bar{X}_n$
  - (See Hansen (2021) section 3.11 for more details on projection and annihilator matrices.)

  then

$$
\begin{aligned}
\frac{n-1}{\sigma^2}S_n^2 &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{\sigma^2}\left((I - P_\iota)X\right)'\left((I - P_\iota)X\right) \\
&= \frac{1}{\sigma^2}X'(I - P_\iota)'(I - P_\iota)X \\
&= \frac{1}{\sigma^2}X'(I - P_\iota)X
\end{aligned}
$$

sample mean and variance

- proof (ii) (cont'd):

$$\frac{1}{\sigma^2}X'(I-P_\iota)X = \frac{1}{\sigma^2}(X-\mu\iota)'(I-P_\iota)(X-\mu\iota)$$

because

$$\begin{aligned}(X-\mu\iota)'(I-P_\iota) &= X'(I-P_\iota)-\mu\iota'(I-P_\iota)\\ &= X'(I-P_\iota)-\mu(\iota'-\iota'P_\iota)\\ &= X'(I-P_\iota)-\mu\left(\iota'-\frac{1}{n}\iota'\iota\iota'\right)\\ &= X'(I-P_\iota)-\mu\left(\iota'-\iota'\right) \quad = \quad X'(I-P_\iota)\end{aligned}$$

so

$$\frac{n-1}{\sigma^2}S_n^2 = \underbrace{\left(\frac{X-\mu\iota}{\sigma}\right)'}_{=Z}(I-P_\iota)\underbrace{\left(\frac{X-\mu\iota}{\sigma}\right)}_{=Z}$$

given that $I-P_\iota$ is a $(n-1)$-dimensional orthogonal projection, it follows from the previous theorem that $\frac{n-1}{\sigma^2}S_n^2 \sim \chi^2_{n-1}$ ■

sample mean and variance

- yet some additional results before proof (iii).
- fact (verify): if $X \sim N(\mu, \Sigma)$ then $AX + B \sim N(A\mu + B, A\Sigma A')$
- theorem: let $Z \sim N(0, I)$ and A and B non-random matrices. Then $A'Z$ and $B'Z$ are independent if, and only if, $A'B = 0$.
- proof: define $C = (A, B)$ and write $CZ \sim N(C\mu, C\Sigma C')$. using the result above, see that the covariance between $A'Z$ and $B'Z$ is zero if, and only if, $A'B = 0$.

independence and chi-squared random variables

- proof (iii): write

$$\bar{X}_n = \frac{1}{n}\iota' P_\iota X$$
$$S_n^2 = \frac{1}{n-1}((I - P_\iota)X)'((I - P_\iota)X)$$

and note that $P_\iota X$ and $(I - P_\iota)X$ are orthogonal:

$$\begin{aligned}(P_\iota X)'(I - P_\iota)X &= X'P_\iota' X - X'P_\iota' P_\iota X \\ &= X'P_\iota X - X'P_\iota X \\ &= 0\end{aligned}$$

hence $P_\iota X$ and $(I - P_\iota)X$ are independent. $\bar{X}_n$ and S_n^2 are functions of independent random variables, so are themselves independent. ■

Student's t distribution

- if $X_1, \ldots, X_n$ be a random sample of independent $N(\mu, \sigma^2)$, then

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \sim N(0,1)$$

- however, most of the time, we do not know σ, and hence the best we can do is to use

$$\begin{aligned} \sqrt{n}\,\frac{\bar{X}_n - \mu}{S_n} &= \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma}\,\frac{1}{\sqrt{S_n^2/\sigma^2}} \\ &= U \cdot \frac{1}{\sqrt{V/(n-1)}} \sim t_{n-1} \end{aligned}$$

given that $\bar{X}_n$ and S_n^2 are independent, $U \sim N(0,1)$ and $V \sim \chi^2_{n-1}$ are independent.

Student's t distribution

- then, since U and V are independent, (to simplify, $p = n - 1$)

$$f_{U,V}(u, v) = \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-u^2/2} \frac{1}{\Gamma\left(\frac{p}{2}\right) 2^{p/2}} v^{(p/2)-1} e^{-v/2}$$

for $-\infty < u < \infty$ and $0 < v < \infty$. Use the transformation

$$t = \frac{u}{\sqrt{v/p}} \quad \text{and} \quad w = v$$

where the inverse functions are

$$u = t\sqrt{w/p} \quad \text{and} \quad v = w$$

with Jacobian

$$J = \begin{vmatrix} \frac{\partial u}{\partial t} & \frac{\partial u}{\partial w} \\ \frac{\partial v}{\partial t} & \frac{\partial v}{\partial w} \end{vmatrix} = \frac{\partial u}{\partial t}\frac{\partial v}{\partial w} - \frac{\partial u}{\partial w}\frac{\partial v}{\partial t} = \sqrt{\frac{w}{p}}$$

Student's t distribution

- So the marginal pdf of T is

$$
\begin{aligned}
f_T(t) &= \int_0^\infty f_{U,V}\left(t\sqrt{\frac{w}{p}}, w\right)\sqrt{\frac{w}{p}}\,\mathrm{d}w \\
&= \frac{1}{(2\pi)^{\frac{1}{2}}}\frac{1}{\Gamma\left(\frac{p}{2}\right)2^{p/2}p^{1/2}}\int_0^\infty e^{-t^2\frac{w}{2p}}w^{(p/2)-1}e^{-w/2}\sqrt{\frac{w}{p}}\,\mathrm{d}w \\
&= \frac{1}{(2\pi)^{\frac{1}{2}}}\frac{1}{\Gamma\left(\frac{p}{2}\right)2^{p/2}p^{1/2}}\underbrace{\int_0^\infty e^{-\frac{w}{2}(1+t^2/p)}w^{((p+1)/2)-1}\,\mathrm{d}w}_{=\text{kernel of } G((p+1)/2,2/(1+t^2/p))} \\
&= \frac{1}{(2\pi)^{\frac{1}{2}}}\frac{1}{\Gamma\left(\frac{p}{2}\right)2^{p/2}p^{1/2}}\Gamma\left(\frac{p+1}{2}\right)\left[\frac{2}{1+t^2/p}\right]^{(p+1)/2}
\end{aligned}
$$

which is the Student's t distribution with parameter p.

- Gamma distribution, $X \sim G(k,\theta)$:

$$
f_X(x) = \frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-\frac{x}{\theta}}
$$

- this completes the proof of (iv)!

Student's t distribution

Some properties of the t distribution:

- with $p = 1$, T_1 becomes the pdf of a Cauchy
- so inference with sample size 2 is impossible!
- $\mathbb{E}(T_p) = 0$ if $p > 1$ and $\mathrm{var}(T_p) = \frac{p}{p-2}$ if $p > 2$
- does not have moments of all orders - no mgf either
- normal distribution approximates well for large p

Snedecor's F distribution

- definition: let $X_1, \ldots, X_n \sim N(\mu_X, \sigma_X^2) \perp\!\!\!\perp Y_1, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$, then

$$\frac{S_{X,n}^2 / S_{Y,m}^2}{\sigma_X^2 / \sigma_Y^2} \;=\; \frac{S_{X,n}^2/\sigma_X^2}{S_{Y,m}^2/\sigma_Y^2} \;=\; \frac{\chi_{n-1}^2/(n-1)}{\chi_{m-1}^2/(m-1)} \;\sim\; F_{n-1,m-1}$$

given that the two chi-squared distributions are independent

$$f(x) \;=\; \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{p/2} \frac{x^{p/2-1}}{[1+(p/q)x]^{(p+q)/2}}$$

with mean $\mathbb{E}(F_{p,q}) = \frac{q}{q-2}$ if $q > 2$, so that the expected value of the variance ratio is approximately one if the sample size is large enough.

- theorem:

  (i) if $X \sim F_{p,q}$, then $1/X \sim F_{q,p}$

  (ii) if $X \sim t_q$, then $X^2 \sim F_{1,q}$

  (iii) if $X \sim F_{p,q}$, then $(p/q)X/(1+(p/q)X) \sim B(p/2, q/2)$

Contents

order statistics

- Some possible applied questions:
  - What is the a maximum rainfall in any given year?
  - The lowest price of a stock?
  - The median value of house prices? (or even quantiles)

- definition: the order statistics of a sample $X_1, \ldots, X_n$ are the sample values placed in ascending order, denoted

$$X_{(1)}, \ldots, X_{(n)}$$

  satisfying $\min_i X_i = X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)} = \max_i X_i$.

- Since X_i are random variables, $X_{(i)}$ are also random variables. Our goal is to describe the pdfs/pmfs for some cases.

order statistics

- of particular interest is the sample median

$$M = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd} \\ \frac{1}{2}X_{(n/2)} + \frac{1}{2}X_{((n+1)/2)} & \text{if } n \text{ is even} \end{cases}$$

which less sensitive to extreme observations (or outliers) than the sample mean.

- the p-quantile is the observation such that np observations are smaller and $n(1-p)$ are greater, $p \in [0,1]$.
  - lower(upper) quartile is the 0.25-quantile (0.75-quantile)

- the sample range,

$$R = X_{(n)} - X_{(1)}$$

which is an alternative measure of dispersion.

order statistics

- theorem (CB 5.4.3): let $X_1, \ldots, X_n$ be a random sample from a discrete distribution with pmf $f_X(x_i) = p_i$, where $x_1 < x_2 < \ldots$ are the possible values of X. Define

$$\begin{aligned} P_0 &= 0 \\ P_1 &= p_1 \\ P_2 &= p_1 + p_2 \\ &\vdots \\ P_i &= \sum_{j=1}^{i} p_i \end{aligned}$$

Then

$$\begin{aligned} \mathbb{P}\left(X_{(j)} \leq x_i\right) &= \sum_{k=j}^{n} \binom{n}{k} P_i^k \left(1 - P_i\right)^{n-k} \\ \mathbb{P}\left(X_{(j)} = x_i\right) &= \sum_{k=j}^{n} \binom{n}{k} \left[P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k}\right] \end{aligned}$$

- proof: fix i and count the number of X_i that are less than or equal to x_i. The event $\{X_j \leq x_i\}$ is a "success", and otherwise a "failure". The question becomes: how many successes Y are there? Given that trials are independent, so $Y \sim \text{Bin}(n, P_i)$.

- the second part only expresses the differences

$$\mathbb{P}\{X_{(j)} = x_i\} \quad = \quad \mathbb{P}\{X_{(j)} \leq x_i\} - P\{X_{(j)} \leq x_{i-1}\} \qquad \blacksquare$$

- there is a similar theorem for the continuous case, but we will do one example instead.

- example: let $X_1, \dots, X_n$ be i.i.d. random variables and define $Y = \max\{X_1, \dots, X_n\}$. The distribution function of Y is given by

$$\begin{aligned} F_Y(y) &= \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \leq y\}\right) \\ &= \prod_{i=1}^{n} \mathbb{P}\{X_i \leq y\} \\ &= (F_Y(y))^n \end{aligned}$$

Contents

Contents

spoiler of next slides

$$\begin{tikzcd}
\xrightarrow{a.s.} \arrow[dr, Rightarrow] & & \\
& \xrightarrow{p} \arrow[r, Rightarrow] & \xrightarrow{d} \\
\xrightarrow{\mathcal{L}^p} \arrow[ur, Rightarrow] & &
\end{tikzcd}$$

non-stochastic convergence

- suppose you have a non-stochastic sequence $\{a_n\}_{n=1}^{\infty}$.
- we say that $\{a_n\}_{n=1}^{\infty}$ converges to a if, and only if, for each $\epsilon > 0$ there exists a $N \in \mathbb{N}$ such that if $n > N$, we have that

$$|a_n - a| \quad < \quad \epsilon$$

  and we write $a_n \longrightarrow a$ or $\lim_{n\to\infty} a_n = a$.

- example 1: $a_n = 1 + \frac{1}{n} \quad \Rightarrow \quad \lim_{n\to\infty} a_n = 1$.
- proof: fix $\epsilon > 0$. We want to select an N such that $|a_n - a| = n^{-1} < \epsilon$ for $n > N$. Set $N = \frac{1}{\epsilon} - 1$. For $n > N = \frac{1}{\epsilon} - 1$, we have that $n^{-1} < \frac{\epsilon}{1-\epsilon} < \epsilon$. So the sequence converges to 1.

non-stochastic convergence

- example 2: $a_n = \frac{sin(n)}{n} \quad \Rightarrow \quad \lim_{n\to\infty} a_n = 0$
- proof: fix $\epsilon > 0$ and choose $N > \frac{1}{\epsilon}$. Since $-1 \leq \sin(n) \leq 1$, we have that $|\sin(n)| < 1$. Therefore

$$\left|\frac{\sin(n)}{n} - 0\right| = \frac{|\sin(n)|}{n} \leq \frac{1}{n} < \frac{1}{N} < \frac{1}{1/\epsilon} = \epsilon$$

- example 3: for any $w \in \mathbb{R}$, define the sequence

$$\{a_1, a_2, \ldots\} = \{w+1, w, w+1, w, w, w+1, w, w, w, w+1, \ldots\}$$

and suggest the limit $a = w$, so $|a_n - a| = \{1, 0, 1, 0, 0, 1, \ldots\}$. If the series converges, for any $\epsilon > 0$, there must exist an N such that $n > N$ implies that $|a_n - a| < \epsilon$.

Take $\epsilon = 2$. It is true that $|a_n - a| < \epsilon$ for any n, so suffices to take $N = 1$.

Take $\epsilon = 0.5$. There isn't an N such that $|a_n - a| < \epsilon$ for every $n > N$, so the sequence does not converge.

non-stochastic convergence

- definition does not apply to sequence of random variables $\{X_n\}$: we would have $|X_n - X| < \epsilon$ sometimes being true, sometimes being false...
- example: take $X_n \sim N(0, \frac{\sigma^2}{n})$ and suggest $X = 0$. Even for "very high" n, it is possible that $|X_n - X| > \epsilon$. So we can never find for sure an N such that $|X_n - X| < \epsilon$ for $n > N$.
- we can only say what is the probability of being true.
- the probability is not a random variable!

convergence in probability

- definition: a sequence of random variables $X_1, X_2, \ldots$ converges in probability to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(|X_n - X| \geq \epsilon) \;=\; 0$$

  or, equivalently,

$$\lim_{n \to \infty} \mathbb{P}(|X_n - X| < \epsilon) \;=\; 1$$

- for reasons that will be clearer soon, we will come back to the σ-algebra notation for an equivalent - and more formal - definition.

- definition: let X_n be defined on a common probability space $(\Omega, \mathcal{F}, P)$. $\{X_n\}$ converges in probability to X if, for any $\epsilon > 0$,

$$\mathbb{P}\left(\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\right) \;\longrightarrow\; 0$$

- if X_n converges in probability to X we write $X_n \xrightarrow{p} X$.

convergence in probability

- example (cont'd): take $X_n \sim N(0, \frac{\sigma^2}{n})$ and suggest $X = 0$.

$$
\begin{aligned}
\mathbb{P}\left(|X_1 - X| < \epsilon\right) &= \Phi(\epsilon/\sigma) - \Phi(-\epsilon/\sigma) = 2 \cdot \Phi(\epsilon/\sigma) - 1 \\
\mathbb{P}\left(|X_2 - X| < \epsilon\right) &= 2 \cdot \Phi(\sqrt{2}\epsilon/\sigma) - 1 \\
&\vdots \\
\mathbb{P}\left(|X_n - X| < \epsilon\right) &= 2 \cdot \Phi(\sqrt{n}\epsilon/\sigma) - 1
\end{aligned}
$$

  where Φ is the cdf of the standard normal.

- From the definition of a cdf, we get that

$$
\lim_{n\to\infty} \Phi(\sqrt{n}\epsilon/\sigma) = 1 \quad \Rightarrow \quad \lim_{n\to\infty} 2 \cdot \Phi(\sqrt{n}\epsilon/\sigma) - 1 = 1
$$

  so the deterministic sequence of probabilities converges to 1, i.e., X_n converges in probability.

convergence in probability

- theorem (**weak law of large numbers**) (CB 5.5.2): let $X_1, X_2, \ldots$ denote iid random variables with $\mathbb{E}(X_i) = \mu$ and $\mathrm{var}(X_i) = \sigma^2 < \infty$, then $\bar{X}_n \xrightarrow{p} \mu$.

- proof: Chebyschev inequality states that

$$\mathbb{P}\left(g(X) \geq r\right) \quad = \quad \frac{1}{r}\mathbb{E}\left[g(X)\right] \ \text{ for any } r > 0$$

and so, selecting $g(X) = |\bar{X}_n - \mu|$ and $r = \epsilon$,

$$\begin{aligned}
\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \quad &= \quad \mathbb{P}\left((\bar{X}_n - \mu)^2 \geq \epsilon^2\right) \\
&\overset{Chebys.}{\leq} \quad \frac{\mathbb{E}\left(\bar{X}_n - \mu\right)^2}{\epsilon^2} \\
&= \quad \frac{\mathrm{var}\left(\bar{X}_n^2\right)}{\epsilon^2} \quad = \quad \frac{\sigma^2}{n\epsilon^2}
\end{aligned}$$

then, for every $\epsilon > 0$, $\frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$. ■

consistency

- if $\hat{\theta}_n$ is a statistic that summarizes the information about θ, then
  (i) $\hat{\theta}_n$ is unbiased if $\mathbb{E}(\hat{\theta}_n) = \theta$
  (ii) $\hat{\theta}_n$ is consistent if $\lim_{n\to\infty} \mathbb{P}(|\hat{\theta}_n - \theta| < \epsilon) = 1$ for every $\epsilon > 0$
- example: showing the consistency of S_n^2 by Chebychev...

$$\mathbb{P}(|S_n^2 - \sigma^2| \geq \epsilon) \;\leq\; \frac{\mathbb{E}(S_n^2 - \sigma^2)}{\epsilon^2} \;=\; \frac{\mathrm{var}(S_n^2)}{\epsilon^2},$$

which converges to zero as long as $\mathrm{var}(S_n^2) \to 0$ as $n \to \infty$ (more on this soon) ■

almost sure convergence

- definition: a sequence of random variables $X_1, X_2, \ldots$ converges almost surely to a random variable X if, for every $\epsilon > 0$,

$$\mathbb{P}\left(\lim_{n\to\infty} |X_n - X| \geq \epsilon\right) \;=\; 0$$

or, equivalently,

$$\mathbb{P}\Big(\omega | X_n(\omega) \to X(\omega)\Big) \;=\; 1$$

- if X_n converges almost surely to X we write $X_n \xrightarrow{a.s.} X$.

- convergence in probability is about the behavior of the sequence as the sample size grows, whereas almost sure convergence is much stronger in that it dictates that $X_n(\omega)$ converges to $X(\omega)$ for all $\omega \in \Omega$, except perhaps for a set of null measure.

almost sure convergence

- example 1: let $\Omega = [0, 1]$ with uniform probability distribution.
- define $X_n(\omega) = \omega^n$ and $X(\omega) = 0$.
- for every $s \in [0, 1)$, $s^n \to 0$ as $n \to \infty$. So, in this subset, $X_n(\omega) \to 0 = X(\omega)$.
- however, $X_n(1) = 1$ for every n, which does not converge to $X(1) = 0$.
- yet, the convergence is "almost" surely since $\mathbb{P}([0, 1)) = 1$, so $X_n \xrightarrow{a.s.} X$.

almost sure convergence

- example 2: let $\Omega = [0, 1]$ with uniform distribution and

$$X_n(\omega) = \frac{1}{n}\omega + \frac{n-1}{n}$$

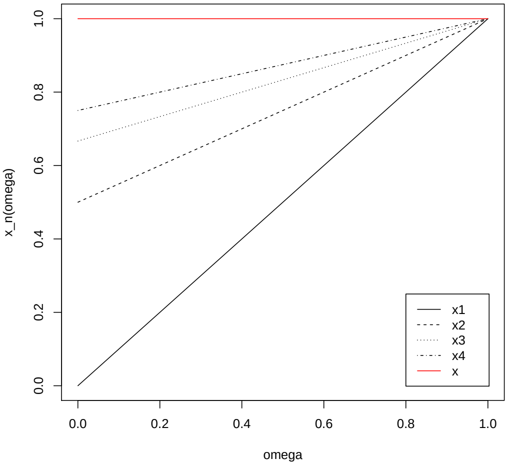
that is,

$$X_1(\omega) = \omega \ ; \ X_2(\omega) = \frac{1}{2}\omega + \frac{1}{2} \ ; \ X_3(\omega) = \frac{1}{3}\omega + \frac{2}{3}$$

and so on.

- We want to check if X_n converges to $X = 1$ in probability and almost surely.

almost sure convergence



- example 2: (almost sure convergence) fix an ω and see if sequence $X_n(\omega)$ converges to $X(\omega)$ as $n \to \infty$. Taking a few values of ω,

$$\begin{aligned} X_n(0.25) &= \{0.25, 0.625, 0.75, 0.8125, 0.85, \ldots, 0.9925, \ldots\} \\ X_n(0.5) &= \{0.5, 0.75, 0.8333, 0.875, 0.9 \ldots, 0.995, \ldots\} \\ X_n(0.75) &= \{0.75, 0.875, 0.9167, 0.9375, 0.95, \ldots, 0.9975, \ldots\} \end{aligned}$$

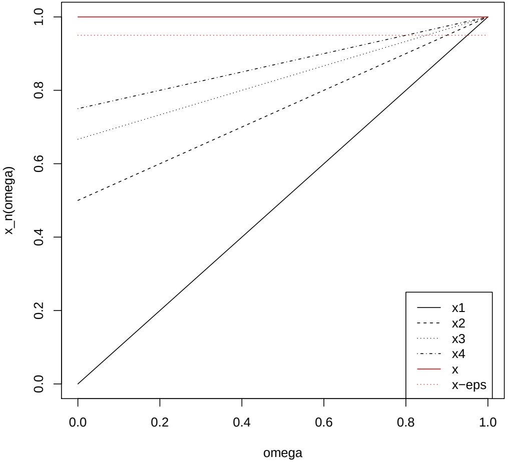
so, for every $\omega \in A = (0, 1]$,

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

and $A^c = \{0\}$. But we have that $\mathbb{P}(A) = 1$, since $\mathbb{P}(\{0\}) = 0$. So $X_n(\omega) \stackrel{a.s.}{\to} X(\omega)$.

convergence in probability

- example 2: (convergence in probability) $\lim_{n\to\infty} \mathbb{P}\left(\omega : |X_n(\omega) - X(\omega)| < \epsilon\right) = 1$



convergence in probability

- example 2: instead of calculating the convergence for a fixed point ω, we look at the convergence of the probability that X_n is not too distant to X.

$$
\begin{aligned}
\mathbb{P}(\omega : |X_1(\omega) - X(\omega)| < \epsilon) &= \mathbb{P}(|\omega - 1| < \epsilon) = \mathbb{P}(-\omega + 1 < \epsilon) \\
&= \mathbb{P}(\omega > 1 - \epsilon) = \epsilon
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}(\omega : |X_2(\omega) - X(\omega)| < \epsilon) &= \mathbb{P}\left(\left|\frac{1}{2}\omega + \frac{1}{2} - 1\right| < \epsilon\right) = \mathbb{P}\left(-\frac{1}{2}\omega + \frac{1}{2} < \epsilon\right) \\
&= \mathbb{P}(\omega > 1 - 2\epsilon) = 2\epsilon
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}(\omega : |X_3(\omega) - X(\omega)| < \epsilon) &= \mathbb{P}\left(\left|\frac{1}{3}\omega + \frac{2}{3} - 1\right| < \epsilon\right) = \mathbb{P}\left(-\frac{1}{3}\omega + \frac{2}{3} < \epsilon\right) \\
&= \mathbb{P}(\omega > 1 - 3\epsilon) = 3\epsilon
\end{aligned}
$$

- the sequence (of probabilities) is $\{\epsilon, 2\epsilon, 3\epsilon, \ldots, 1, 1 \ldots\}$ which converges to 1. So, X_n converges in probability to X, or $X_n \xrightarrow{p} X$.

almost sure convergence

- example 3: let, again, $\Omega = [0,1]$ with uniform distribution and define

$$X_1(\omega) \; = \; \omega + I_{[0,1]}(\omega)$$
$$X_2(\omega) \; = \; \omega + I_{[0,\frac{1}{2}]}(\omega) \; , \; X_3(\omega) \; = \; \omega + I_{(\frac{1}{2},1]}(\omega)$$
$$X_4(\omega) \; = \; \omega + I_{[0,\frac{1}{3}]}(\omega) \; , \; X_5(\omega) \; = \; \omega + I_{(\frac{1}{3},\frac{2}{3}]}(\omega) \; , \; X_6(\omega) \; = \; \omega + I_{(\frac{2}{3},1]}(\omega)$$

  and $X(\omega) = \omega$.

- $\mathbb{P}(|X_n - X| \geq \epsilon) = (b_n - a_n)$, where $\lim_{n\to\infty}(b_n - a_n) = 0$. So $X_n \xrightarrow{p} X$.

- however, there is no value $\omega \in \Omega$ such that $X_n(\omega) \to \omega = X(\omega)$.

- to see this, fix any $\omega \in \Omega$. As n grows, we will see a sequence of the type

$$\omega + 1, \omega, \omega + 1, \omega, \omega + 1, \omega, \ldots$$

  in which $\omega + 1$ appear infinitely often. Therefore $X_n(\omega) \overset{a.s}{\not\to} X$.

strong law of large numbers

- theorem (**strong law of large numbers**) (CB 5.5.9): let $X_1, X_2, \ldots$ denote a sequence of iid random variables such that $\mathbb{E}(X_i) = \mu$ and $\text{var}(X_i) = \sigma^2 < \infty$, then

$$\mathbb{P}\left(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon\right) = 1 \qquad \text{for every } \epsilon > 0$$

that is, $\bar{X}_n \xrightarrow{a.s.} \mu$.

relation between modes of convergence

- the counterexample shows that $X_n \xrightarrow{p} X \quad \not\Rightarrow \quad X_n \xrightarrow{a.s.} X$.

- theorem: $X_n \xrightarrow{a.s.} X \quad \Rightarrow \quad X_n \xrightarrow{p} X$

- proof: consider a sequence of events

$$S_n \quad = \quad \bigcup_{m \geq n} \{\omega : |X_m(\omega) - X(\omega)| > \epsilon\}$$

and so

$$\mathbb{P}\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\} \quad \leq \quad \mathbb{P}\{S_n\}.$$

Note that $S_n \supseteq S_{n+1} \supseteq S_{n+2} \supseteq \cdots$ and, in the limit, decreases towards

$$S_\infty \quad = \quad \bigcap_{n \geq 1} S_n$$

with $\mathbb{P}\{S_n\} \xrightarrow{n \to \infty} \mathbb{P}\{S_\infty\}$. We will show that if $X_n \xrightarrow{a.s.} X$, then $\mathbb{P}\{S_\infty\} = 0$.

- proof (cont'd): if $X_n \xrightarrow{a.s.} X$, then the set

$$S_0 \quad = \quad \left\{ \omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega) \right\}$$

is such that $\mathbb{P}(S_0) = 0$. So if we show that $S_\infty \subseteq S_0$, then $\mathbb{P}(S_\infty) = 0$ and we're done.

Take any point $\omega \notin S_0$. It is such that for a certain $n \geq N$

$$\lim_{n\to\infty} X_n(\omega) \quad = \quad X(\omega) \quad \Longrightarrow \quad |X_n(\omega) - X(\omega)| \quad < \quad \epsilon.$$

This implies that for $n \geq N$, $\omega \notin S_n$ and so $\omega \notin S_\infty$. This means that $S_\infty \subseteq S_0$. ■

convergence in distribution

- definition: a sequence of random variables $X_1, X_2, \ldots$ converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) \;=\; F_X(x)$$

  at all points x in which $F_X(x)$ is continuous.

- example (CB 5.5.11): if $X_1, X_2, \ldots$ are iid U(0,1) and $X_{(n)} = \max_{1\le i\le n} X_i$, then we expect $X_{(n)}$ to approach one from below

$$\begin{aligned}
\mathbb{P}(|X_{(n)} - 1| \geq \epsilon) &= \mathbb{P}(X_{(n)} \geq 1+\epsilon) + \mathbb{P}(X_{(n)} \leq 1-\epsilon) \\
&= \mathbb{P}(X_{(n)} \leq 1-\epsilon) \\
&= \mathbb{P}(X_i \leq 1-\epsilon \text{ for } i = 1,\ldots,n) \\
&= \left[\mathbb{P}(X_i \leq 1-\epsilon)\right]^n \\
&= (1-\epsilon)^n \to 0
\end{aligned}$$

  (so $X_{(n)}$ converges to 1 in probability).

- example (cont'd): Taking $\epsilon = t/n$,

$$\mathbb{P}\left(X_{(n)} \leq 1-\epsilon\right) \;=\; \mathbb{P}\left(X_{(n)} \leq 1-t/n\right) \;=\; (1-t/n)^n \;=\; e^{-t}$$

since $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$. Upon rearranging,

$$\mathbb{P}\left(n(1-X_{(n)}) \leq t\right) \;=\; 1-e^{-t}$$

and the random variable $n\left(1-X_{(n)}\right)$ converges in distribution to an exp(1). ■

- it is really about convergence of the cdfs, not the random variables

relation between modes of convergence

- theorem: the following are equivalent:

  (i) $X_n \xrightarrow{d} X$

  (ii) $F_n(t) \to F(t)$ for every continuity point of F

  (iii) $\mathbb{E}g(X_n) \longrightarrow \mathbb{E}g(X)$ for every bounded and uniformly continuous g

- definition: a function g is uniformly continuous if for every real number $\varepsilon > 0$ there exists a real number $\delta > 0$ such that for every $x, y \in X$, $|x - y| < \delta \implies |g(x) - g(y)| < \varepsilon$.

relation between modes of convergence

- theorem (CB 5.5.12): $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$
- proof: pick an arbitrary g that is bounded and uniformly continuous and let $M = \sup |g(x)|$. For any $\epsilon > 0$, choose δ such that

$$|X_n - X| \leq \delta \;\Rightarrow\; |g(X_n) - g(X)| \leq \epsilon$$

we have that

$$|g(X_n) - g(X)| \leq \epsilon I\{|X_n - X| \leq \delta\} + 2M \cdot I\{|X_n - X| > \delta\}$$

it follows that

$$\begin{aligned} |\mathbb{E}g(X_n) - \mathbb{E}g(X)| &\leq \mathbb{E}|g(X_n) - g(X)| \\ &\leq \epsilon \mathbb{P}\{|X_n - X| \leq \delta\} + 2M \cdot \mathbb{P}\{|X_n - X| > \delta\} \end{aligned}$$

because $|\mathbb{E}(W)| \leq \mathbb{E}(|W|)$ for any W. The conclusion follows. ■

- corollary: $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{d} X$

relation between modes of convergence

- theorem (CB 5.5.13): X_n converges in probability to a constant μ if, and only if, the sequence also converges in distribution to μ. That is,

$$\mathbb{P}\left(|X_n - \mu| > \epsilon\right) \quad \rightarrow \quad 0 \quad \text{for every } \epsilon > 0$$

is equivalent to

$$\mathbb{P}\left(X_n \leq x\right) \quad \rightarrow \quad \begin{cases} 0 & \text{if } x < \mu \\ 1 & \text{if } x \geq \mu \end{cases}$$

- proof: Casella & Berger, exercise 5.41.

central limit theorem

- theorem (**central limit theorem**) (CB 5.5.14): let $X_1, X_2, \ldots$ denote a sequence of iid random variables whose mgfs exist in a neighborhood of zero, that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h. Let $\mathbb{E}(X_i) = \mu$ and $\text{var}(X_i) = \sigma^2 > 0$, both finite. Then the cdf $G_n(x)$ of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ is such that

$$\lim_{n\to\infty} G_n(x) \;=\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u$$

  and hence $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$

- kind of magic: virtually no assumptions and we end up with normality!
- some intuition: sums of “small” (finite variance) independent disturbances
  - example: a Cauchy variable will not converge to a Normal

CLT proof

- proof: let $Z_i = \frac{X_i - \mu}{\sigma}$, then $\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i = \sqrt{n}\frac{(\bar{X}_n - \mu)}{\sigma}$. By properties of mgfs,

$$M_{\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}}(t) = M_{\sum_{i=1}^n Z_i/\sqrt{n}}(t) = M_{\sum_{i=1}^n Z_i}(t/\sqrt{n}) = \left[M_Z(t/\sqrt{n})\right]^n$$

expanding $M_Z(t/\sqrt{n})$ into a Taylor series, we get

$$\begin{aligned}\left[M_Z(t/\sqrt{n})\right]^n = \mathbb{E}\left(e^{\frac{t}{\sqrt{n}}X}\right)^n &= \left[\mathbb{E}\left(\sum_{k=0}^{\infty} X^n \frac{(t/\sqrt{n})^k}{k!}\right)\right]^n \\ &= \left[\sum_{k=0}^{\infty} M_Z^{(k)}(0)\frac{(t/\sqrt{n})^k}{k!}\right]^n\end{aligned}$$

where $M_Z^{(k)}(0) = \frac{d^k}{dt^k} M_Z(t)\Big|_{t=0}$. Since the mgfs exists for $|t| < h$, the Taylor expansion exists for $|t| < h$. Using that, by construction, $M_Z^{(0)}(0) = 1$, $M_Z^{(1)}(0) = 0$ and $M_Z^{(2)}(0) = 1$,

$$\left[\sum_{k=0}^{\infty} M_Z^{(k)}(0)\frac{(t/\sqrt{n})^k}{k!}\right]^n = \left[1 + \frac{(t/\sqrt{n})^2}{2} + R_Z(t/\sqrt{n})\right]^n$$

CLT proof

- proof (cont'd): therefore

$$\lim_{n\to\infty} M_{\sqrt{n}\,\frac{\bar{X}_n-\mu}{\sigma}}(t) \quad = \quad \lim_{n\to\infty}\left[1+\frac{1}{n}\frac{t^2}{2}+R_Z\left(t/\sqrt{n}\right)\right]^n$$

using the facts that $\lim_{n\to\infty} R_Z(t/\sqrt{n}) = 0$ (Taylor's theorem) and

$$\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n \quad = \quad e^a$$

where $a = \lim_{n\to\infty} a_n$, we get that

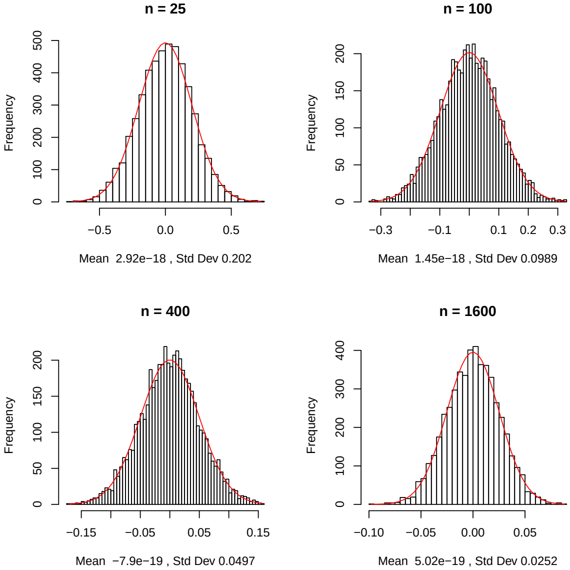
$$\lim_{n\to\infty} M_{\sqrt{n}\,\frac{\bar{X}_n-\mu}{\sigma}}(t) \quad = \quad e^{(t^2/2)}$$

which is the mgf of a standard normal! ☺ ■

CLT in practice + R codes

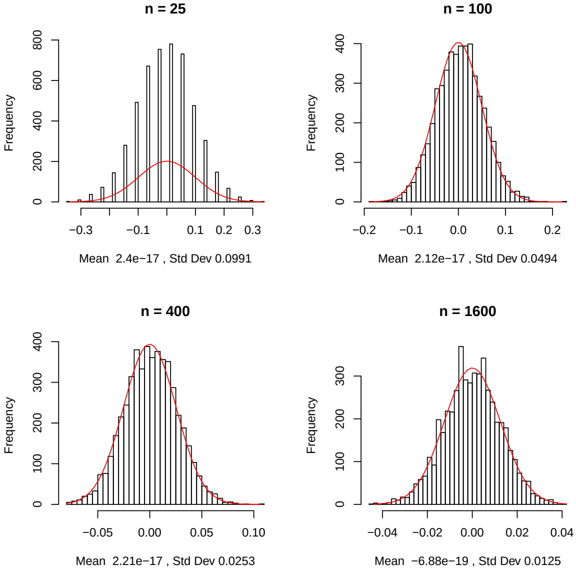
```
samplerCLT <- function(n,choice){
 x <- matrix(0,5000,1)
 for (i in 1:5000){
  if (choice == 1){x[i] <- mean(rnorm(n))}
  if (choice == 2){x[i] <- mean(rbinom(n,1,0.5))}
  if (choice == 3){x[i] <- mean(rbinom(n,1,0.05))}
  if (choice == 4){x[i] <- mean(rbinom(n,10,0.3))}
  if (choice == 5){x[i] <- mean(rbeta(n,2,3))}
  if (choice == 6){x[i] <- mean(rchisq(n,3))}
  if (choice == 7){x[i] <- mean(rt(n,3))}
  if (choice == 8){x[i] <- mean(rcauchy(n))}
  if (choice == 9){x[i] <- mean(rlnorm(n))}
 }
 x <- (x-mean(x))
 h <- hist(x,breaks=50,main=paste('n =',toString(n)),xlab=paste('Mean '
  ,format(mean(x),digits=3),', Std Dev',format(sd(x),digits=3)))
 xfit <- seq(min(x),max(x),length=50)
 yfit <- dnorm(xfit,mean=mean(x),sd=sd(x))
 yfit <- yfit*diff(h$mids[1:2])*length(x)
 lines(xfit,yfit,col='red')
}
```

std normal

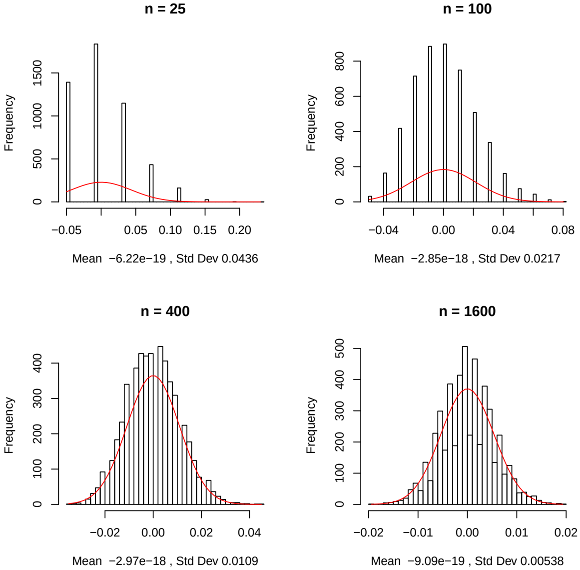


n = 25
Frequency
0 100 200 300 400 500
−0.5 0.0 0.5
Mean 2.92e−18 , Std Dev 0.202
n = 100
Frequency
0 50 100 150 200
−0.3 −0.1 0.1 0.2 0.3
Mean 1.45e−18 , Std Dev 0.0989
n = 400
Frequency
0 50 100 150 200
−0.15 −0.05 0.05 0.15
Mean −7.9e−19 , Std Dev 0.0497
n = 1600
Frequency
0 100 200 300 400
−0.10 −0.05 0.00 0.05
Mean 5.02e−19 , Std Dev 0.0252

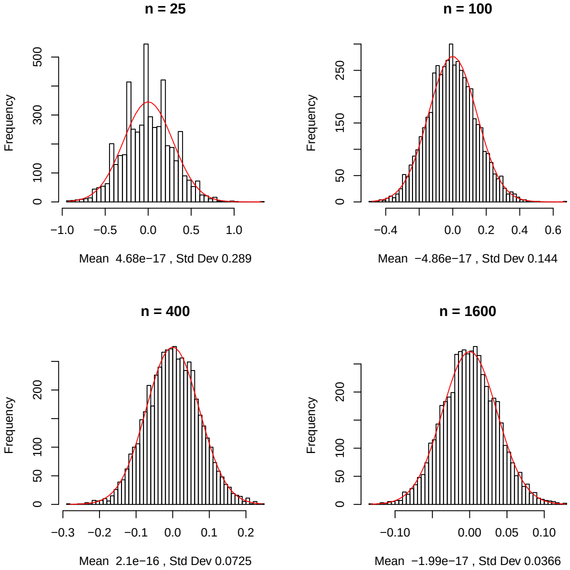
bernoulli(0.5)



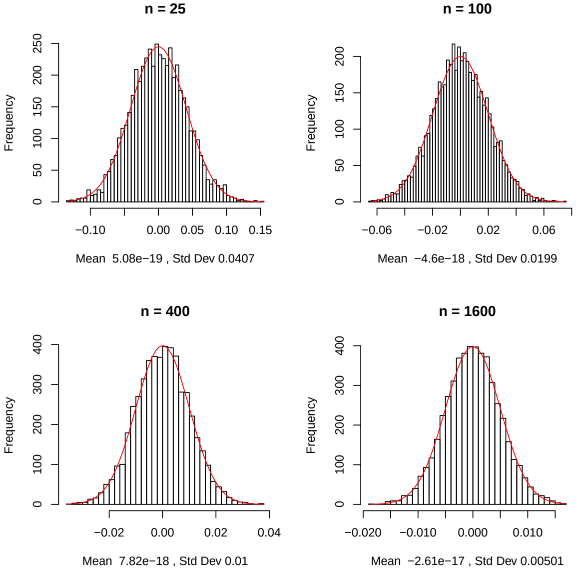
bernoulli(0.05)



binomial(10,0.3)

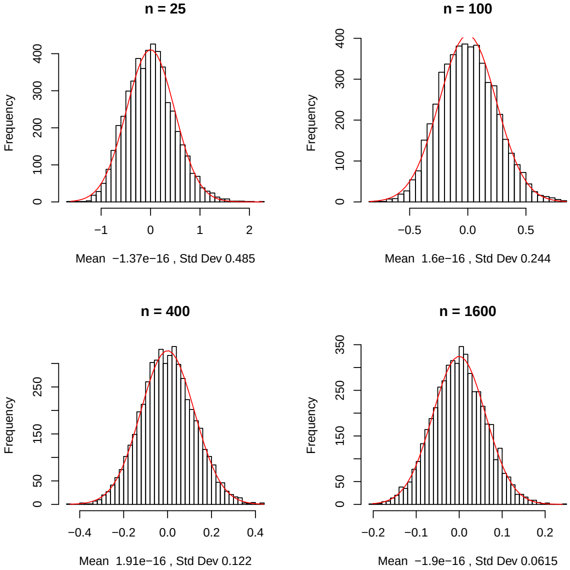


beta(2,3)

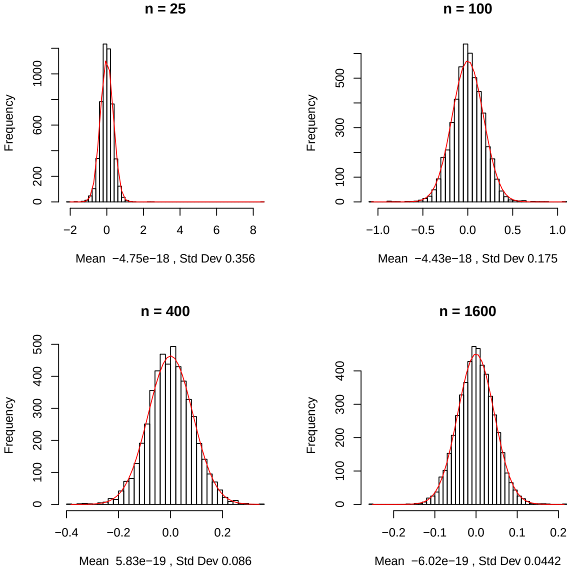


n = 25
Frequency
0
50
100
150
200
250
−0.10
0.00
0.05
0.10
0.15
Mean 5.08e−19 , Std Dev 0.0407
n = 100
Frequency
0
50
100
150
200
−0.06
−0.02
0.02
0.06
Mean −4.6e−18 , Std Dev 0.0199
n = 400
Frequency
0
100
200
300
400
−0.02
0.00
0.02
0.04
Mean 7.82e−18 , Std Dev 0.01
n = 1600
Frequency
0
100
200
300
400
−0.020
−0.010
0.000
0.010
Mean −2.61e−17 , Std Dev 0.00501

chi-squared(3)

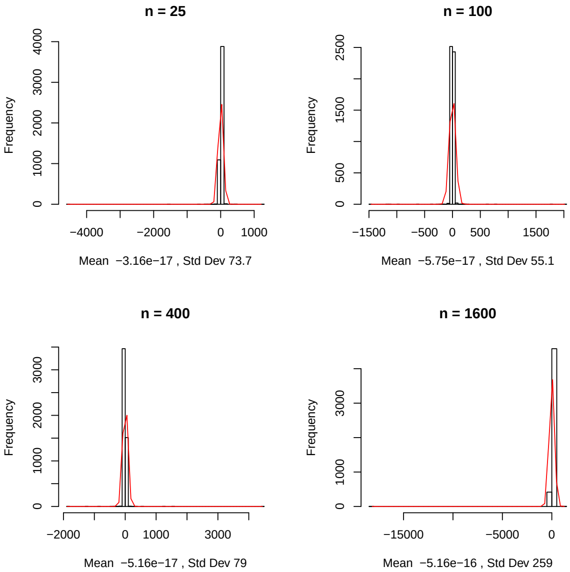


t(3)



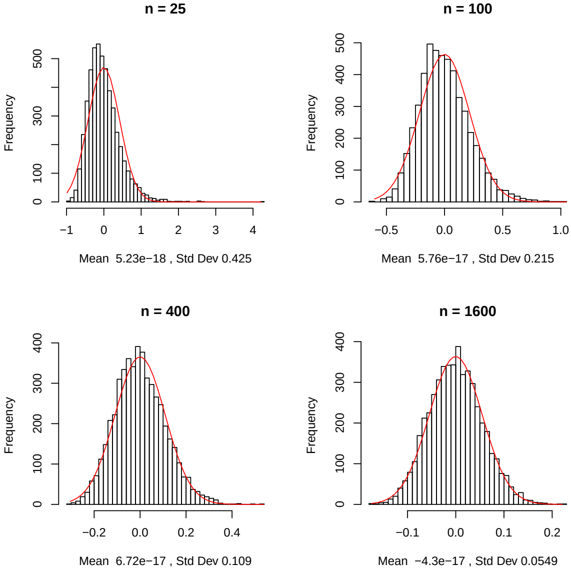
n = 25
Frequency
0
200
600
1000
−2
0
2
4
6
8
Mean −4.75e−18 , Std Dev 0.356
n = 100
Frequency
0
100
300
500
−1.0
−0.5
0.0
0.5
1.0
Mean −4.43e−18 , Std Dev 0.175
n = 400
Frequency
0
100
200
300
400
500
−0.4
−0.2
0.0
0.2
Mean 5.83e−19 , Std Dev 0.086
n = 1600
Frequency
0
100
200
300
400
−0.2
−0.1
0.0
0.1
0.2
Mean −6.02e−19 , Std Dev 0.0442

cauchy



n = 25
Frequency
0
1000
2000
3000
4000
−4000
−2000
0
1000
Mean −3.16e−17 , Std Dev 73.7
n = 100
Frequency
0
500
1500
2500
−1500
−500
0
500
1500
Mean −5.75e−17 , Std Dev 55.1
n = 400
Frequency
0
1000
2000
3000
−2000
0
1000
3000
Mean −5.16e−17 , Std Dev 79
n = 1600
Frequency
0
1000
3000
−15000
−5000
0
Mean −5.16e−16 , Std Dev 259

log-normal



n = 25
Frequency
0
100
300
500
−1
0
1
2
3
4
Mean 5.23e−18 , Std Dev 0.425
n = 100
Frequency
0
100
200
300
400
500
−0.5
0.0
0.5
1.0
Mean 5.76e−17 , Std Dev 0.215
n = 400
Frequency
0
100
200
300
400
−0.2
0.0
0.2
0.4
Mean 6.72e−17 , Std Dev 0.109
n = 1600
Frequency
0
100
200
300
400
−0.1
0.0
0.1
0.2
Mean −4.3e−17 , Std Dev 0.0549

$\mathcal{L}^p$ and moments convergence

- definition: we say that X_n converges in $\mathcal{L}^p$ ("norm-p") to X if

$$\mathbb{E}|X_n - X|^p \quad \longrightarrow \quad 0$$

  and we write $X_n \xrightarrow{\mathcal{L}^p} X$.

- a common particular case is taking $p = 2$, the $\mathcal{L}^2$-convergence, also known as mean squared error convergence.

- definition: we say that X_n converges to X in the p-th moments if

$$\mathbb{E}X_n^p \quad \longrightarrow \quad \mathbb{E}X$$

  which is a fairly weak mode of convergence.

$\mathcal{L}^p$ and moments convergence

- theorem: $X_n \xrightarrow{\mathcal{L}^p} X \Rightarrow X_n \xrightarrow{p} X$

- proof: it is an immediate application of Chebyschev inequality, since

$$\mathbb{P}\left(|X_n - X| > \epsilon\right) \quad \leq \quad \frac{\mathbb{E}|X_n - X|^p}{\epsilon^p}$$

  which completes the proof. ■

- so we have the following summary scheme:

$$
\begin{array}{ccccccc}
\xrightarrow{a.s.} & & & & & & \\
& \searrow & & & & & \\
& & \xrightarrow{p} & \Longrightarrow & \xrightarrow{d} & & \\
& \nearrow & & & & & \\
\xrightarrow{\mathcal{L}^p} & & & & & &
\end{array}
$$

Contents

stochastic order

- we introduce a set of notations known as Mann and Wald's $O_p(1)$ and $o_p(1)$.
- definition: let $\{a_n\}$ and $\{b_n\}$ be a sequence of deterministic real numbers. We write $x_n = o(a_n)$ and $y_n = O(b_n)$ if

$$\frac{x_n}{a_n} \rightarrow 0 \quad \text{and} \quad \left|\frac{y_n}{b_n}\right| < M$$

  $n > N$ and for $M > 0$.

- in particular,
  - $x_n = o(1)$ if x_n converges to zero
  - $y_n = O(1)$ if the sequence is bounded
  - sequence is bounded if converges to zero, so $o(1) = O(1)$
- the sign "=" is not really an equality: $o(1) = O(1)$ but $O(1) \neq o(1)$
  - read as the verb "to be": $o(1)$ *is* $O(1)$, and $O(1)$ *is not* $o(1)$

stochastic order

- example 1: $x_n = 1 + \frac{1}{n} \to 1$, so $x_n \neq o(1)$ but $x_n = O(1)$ and $x_n = o(n)$
- example 2: $x_n = \frac{sin(n)}{n} \to 0$, so $x_n = o(1)$
- example 3: $x_n = \{1, 0, 1, 0, 0, 1, \ldots\}$, so $x_n \neq o(1)$ but $x_n = O(1)$ and $x_n = o(n)$
- example 4: $x_n = n^2$, $x_n \neq o(1)$, $x_n \neq o(n)$, $x_n \neq o(n^2)$, $x_n = o(n^3)$, $x_n \neq O(1)$, $x_n \neq O(n)$, $x_n = O(n^2)$
- example 5: say that $x_n = o(a_n)$. Then $x_n = a_n \frac{x_n}{a_n} = a_n o(1)$
- example 6: say that $y_n = o(b_n)$. Then $y_n = b_n \frac{y_n}{b_n} = b_n O(1)$

stochastic order

- theorem: we have
  (i) $O(o(1)) = o(1)$
  (ii) $o(O(1)) = o(1)$
  (iii) $o(1)O(1) = o(1)$

- proof (i): let $x_n = o(1)$ and $y_n = O(x_n)$. With M such that $\left|\frac{y_n}{x_n}\right| < M$, we have that

$$|y_n| \quad < \quad M|x_n| \quad \to \quad 0$$

- proof (ii): assume that $x_n = O(1)$ and $y_n = o(x_n)$. Choose M such that $|x_n| < M$. Then it follows that

$$\frac{|y_n|}{M} \quad < \quad \left|\frac{y_n}{x_n}\right| \quad \to \quad 0$$

- proof (iii): let $x_n = o(1)$ and $y_n = O(1)$. Then, for M such that $|y_n| < M$,

$$|x_n y_n| \quad < \quad |x_n|M \quad \to \quad 0$$

stochastic order

- definition: we write $X_n = o_p(a_n)$ if X_n/a_n converges in probability to zero.
- in particular, $X_n = o_p(1)$ if $X_n \xrightarrow{p} 0$.
- definition: we write $Y_n = O_p(b_n)$ if for any $\epsilon > 0$, there exists $M > 0$ such that

$$\mathbb{P}\left\{\left|\frac{Y_n}{b_n}\right| > M\right\} \quad < \quad \epsilon$$

  and when $Y_n = O_p(1)$, there exists M such that $\mathbb{P}\{|Y_n| < M\} < \epsilon$ for any $\epsilon > 0$. We then say that Y_n is stochastically bounded.

- we also have that $O_p(o_p(1)) = o_p(O_p(1)) = o_p(1)O_p(1) = o_p(1)$

- theorem: if $X_n \xrightarrow{d} X$, and $Y_n \xrightarrow{p} c$ then
  (i) $X_n = O_p(1)$
  (ii) $X_n + o_p(1) \xrightarrow{d} X$
  (iii) $X_n Y_n \xrightarrow{d} cX$

- proof (i): fix $\epsilon > 0$ and choose M such that $\mathbb{P}(|X| > M) < \epsilon$. Since $X_n \xrightarrow{d} X$, there exists some N such that for $n > N$, we have

$$\mathbb{P}\left(|X| > M\right) < \epsilon I\left\{|X_n - X| \le \delta\right\} \quad \Rightarrow \quad \mathbb{P}\left(|X_n| > M\right) < \epsilon.$$

- proof (ii): let $Y_n = o_p(1)$ and assume that f is uniformly continuous and bounded. It suffices to show that the following term approaches 0.

$$\begin{aligned} |\mathbb{E}f(X_n + Y_n) - \mathbb{E}f(X)| &\leq |\mathbb{E}f(X_n + Y_n) - \mathbb{E}f(X_n)| + |\mathbb{E}f(X_n) - \mathbb{E}f(X)| \\ &\leq \mathbb{E}|f(X_n + Y_n) - f(X_n)| + |\mathbb{E}f(X_n) - \mathbb{E}f(X)| \end{aligned}$$

  the second term is arbitrarily small since $X_n \xrightarrow{d} X$. The first term is also arbitrarily small since

$$|f(X_n + Y_n) - f(X_n)| \leq \epsilon \cdot I\{|Y_n| \leq \delta\} + 2M \cdot I\{|Y_n| > \delta\}$$

  where $M = \sup |f(x)|$ and ϵ and δ are chosen as in the proof where we showed that $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$.

- The final step follows from taking expectations on both sides and noticing that $\mathbb{P}\{|Y_n| > \delta\} \to 0$.■

- proof (iii): since $X_n = O_p(1)$ and $Y_n = c + o_p(1)$,

$$\begin{aligned} X_n Y_n &= X_n(c + o_p(1)) \\ &= cX_n + O_p(1)o_p(1) \\ &= cX_n + o_p(1) \end{aligned}$$

  and the apply (ii) ∎

- Some remarks:
  - let $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$. Then $X_n + Y_n \xrightarrow{p} X + Y$.
  - however, $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ does not imply that $X_n + Y_n \xrightarrow{d} X + Y$, since the joint distribution needs to be taken into consideration!
  - by the CLT, we have that $\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} N(0,1)$. So $\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} = O_p(1)$. We may equivalently say that $\frac{\bar{X}_n - \mu}{\sigma} = O_p\left(1/\sqrt{n}\right)$, or $\bar{X}_n = O_p\left(1/\sqrt{n}\right)$, or $\bar{X}_n - \mu = o_p(1)$.

- theorem (**Slutsky**) (CB 5.5.17): if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, then
  (i) $X_n Y_n \xrightarrow{d} aX$
  (ii) $X_n + Y_n \xrightarrow{d} X + a$

- typical application: suppose that the CLT holds and hence

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} N(0,1)$$

if σ is unknown, then we may employ a consistent estimator, say S_n,

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{S_n} = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma}\,\frac{\sigma}{S_n} \xrightarrow{d} N(0,1)$$

given that the first fraction converges in distribution to a standard normal distribution, whereas the second fraction converges to one in probability.

CMT and beyond

- theorem: let $h(\cdot)$ be a continuous function

  (i) $X_n \xrightarrow{a.s.} X \Rightarrow h(X_n) \xrightarrow{a.s.} h(X)$

  (ii) $X_n \xrightarrow{p} X \Rightarrow h(X_n) \xrightarrow{p} h(X)$

  (iii) $X_n \xrightarrow{d} X \Rightarrow h(X_n) \xrightarrow{d} h(X)$ (**continuous mapping theorem**)

- theorem (**Cramer-Wold device**): let $\{X_n\}$ be a sequence of random vector. Then

$$X_n \xrightarrow{d} X \quad \Longleftrightarrow \quad \lambda' X_n \xrightarrow{d} \lambda X$$

Contents

the delta method

- the CLT shows that under fairly general conditions a standardized random variable has a limit normal distribution. However, we are often interested in the distribution of functions of this random variable.
- example 1: what is the distribution of $\bar{X}_n^2$ as $n \to \infty$?
- example 2: what is the distribution of $\exp(\bar{X}_n)$ as $n \to \infty$?
- example 3: Brazil and Germany play n matches and the results are $\{X_1, X_2, \ldots, X_n\}$ with $X_i \sim$Bernoulli(p), where p is the probability that Brazil wins. We may estimate $\hat{p} = \bar{X}_n$. However, betting agencies use the odds $\frac{p}{1-p}$, so we might consider estimating the odds by $\frac{\hat{p}}{1-\hat{p}}$. But what are the properties of this estimator?

the delta method

- theorem (**delta method**) (CB 5.5.24): let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$. For a given function g and specific value θ, suppose that $g'(\theta)$ exists and is not 0. Then

$$\sqrt{n}\left[g(Y_n) - g(\theta)\right] \quad \xrightarrow{d} \quad n\left(0, \sigma^2[g'(\theta)]^2\right)$$

ps: (CB ex. 5.43) if $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$, then $Y_n \xrightarrow{p} \theta$

- proof: performing a first-order Taylor expansion,

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

where $R(Y_n, \theta) \to 0$ as $Y_n \to \theta$. Since $Y_n \xrightarrow{p} \theta$ it follows that $R(Y_n, \theta) \xrightarrow{p} 0$. Apply the Slutsky theorem to

$$\sqrt{n}\left[g(Y_n) - g(\theta)\right] \quad = \quad g'(\theta)\sqrt{n}(Y_n - \theta)$$

and the result follows. ∎

the delta method

- example 1 (cont'd): from the CLT,

$$\sqrt{n}(\bar{X}_n - \mu) \quad \xrightarrow{d} \quad N(0, \sigma^2)$$

so, from the delta method, using $g(x) = x^2 \Rightarrow g'(x) = 2x \Rightarrow g'(\mu) = 2\mu$,

$$\sqrt{n}(\bar{X}_n^2 - \mu^2) \quad \xrightarrow{d} \quad N(0, (2\mu)^2\sigma^2)$$

note, however, that $\mu \neq 0$ or the distribution is degenerate.

- example 2 (cont'd): we should use $g(x) = \exp(x) \Rightarrow g'(\mu) = \exp(\mu)$ so

$$\sqrt{n}(\exp(\bar{X}_n) - \exp(\mu)) \quad \xrightarrow{d} \quad N(0, (\exp(\mu))^2\sigma^2)$$

the delta method

- example 3 (cont'd): by the CLT, we have that

$$\sqrt{n}(\hat{p} - p) \quad \xrightarrow{d} \quad N(0, p(1-p))$$

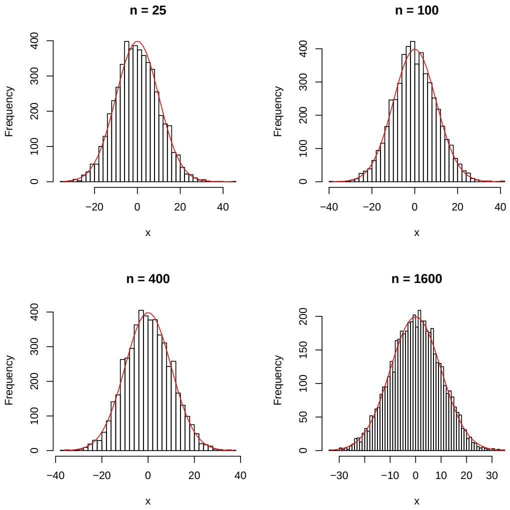
take $g(p) = \frac{p}{1-p}$, so $g'(p) = \frac{1}{(1-p)^2}$ and

$$\begin{aligned}
\sqrt{n}\left(\frac{\hat{p}}{1-\hat{p}} - \frac{p}{1-p}\right) \quad & \xrightarrow{d} \quad N\left(0, \left[g'(p)\right]^2 p(1-p)\right) \\
& \xrightarrow{d} \quad N\left(\left[\frac{1}{(1-p)^2}\right]^2 p(1-p)\right) \\
& \xrightarrow{d} \quad N\left(0, \frac{p}{(1-p)^3}\right)
\end{aligned}$$

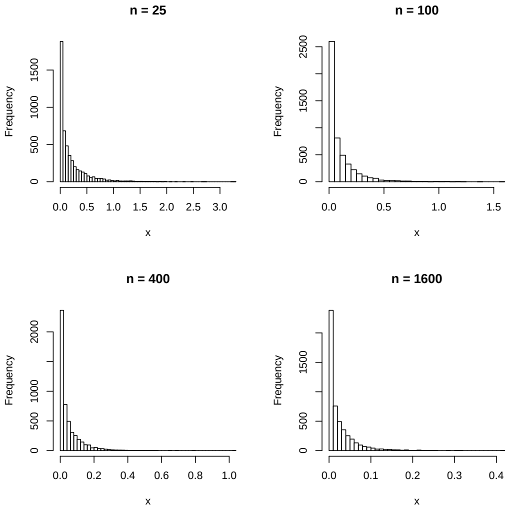
the delta method in practice: example 1

```
samplerDeltaMethodEx1 <- function(n,mu,sigma){

 x <- matrix(0,5000,1)

 for (i in 1:5000){x[i] <- (mean(rnorm(n,mu,sigma)))^2}

 x <- sqrt(n)*(x-(mu)^ 2)
 h <- hist(x,breaks=50,main=paste('n =',toString(n)))

 if (mu!= 0){

   xfit <- seq(min(x),max(x),length=50)
   yfit <- dnorm(xfit,mean=0,sd=2*mu*sigma)
   yfit <- yfit*diff(h$mids[1:2])*length(x)
   lines(xfit,yfit,col='red')

 }

}
```

normal(5,1)


n = 25
n = 100
n = 400
n = 1600
Frequency
x
0
100
200
300
400
−20
20
40
−40
150
50
−30
−10
10
30

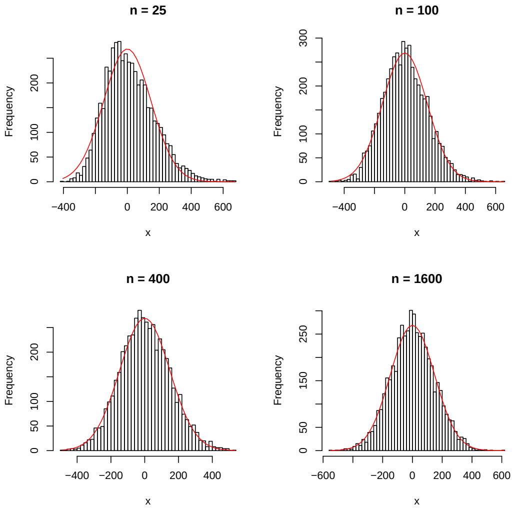
normal(0,1)


n = 25
Frequency
0
500
1000
1500
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x
n = 100
Frequency
0
500
1500
2500
0.0
0.5
1.0
1.5
x
n = 400
Frequency
0
500
1000
2000
0.0
0.2
0.4
0.6
0.8
1.0
x
n = 1600
Frequency
0
500
1500
0.0
0.1
0.2
0.3
0.4
x

the delta method in practice: example 2

```
samplerDeltaMethodEx2 <- function(n,mu,sigma){

 x <- matrix(0,5000,1)

 for (i in 1:5000){x[i] <- exp(mean(rnorm(n,mu,sigma)))}

 x <- sqrt(n)*(x-exp(mu))
 h <- hist(x,breaks=50,main=paste('n =',toString(n)))

 xfit <- seq(min(x),max(x),length=50)
 yfit <- dnorm(xfit,mean=0,sd=exp(mu)*sigma)
 yfit <- yfit*diff(h$mids[1:2])*length(x)
 lines(xfit,yfit,col='red')

}
```

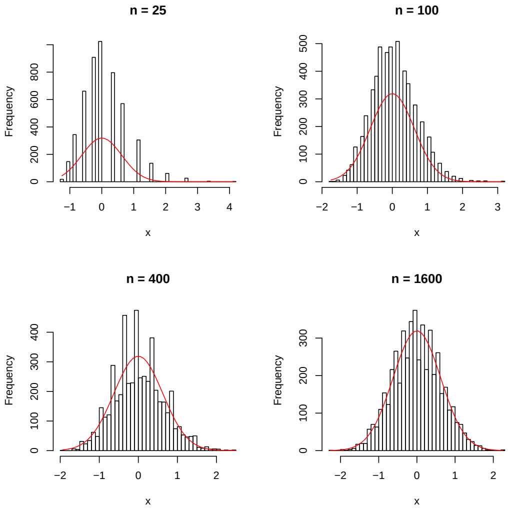
normal(5,1)


n = 25
Frequency
0
50
100
200
−400
0
200
400
600
x
n = 100
Frequency
0
50
100
200
300
−400
0
200
400
600
x
n = 400
Frequency
0
50
100
200
−400
−200
0
200
400
x
n = 1600
Frequency
0
50
150
250
−600
−200
0
200
400
600
x

the delta method in practice: example 3

```
samplerDeltaMethodEx3 <- function(n,mu){

 x <- matrix(0,5000,1)

 for (i in 1:5000){

   phat <- mean(rbinom(n,1,mu))
   x[i] <- phat/(1-phat)

 }

 x <- sqrt(n)*(x-mu/(1-mu))
 h <- hist(x,breaks=50,main=paste('n =',toString(n)))

 xfit <- seq(min(x),max(x),length=50)
 yfit <- dnorm(xfit,mean=0,sd=sqrt(mu/((1-mu)^3)))
 yfit <- yfit*diff(h$mids[1:2])*length(x)
 lines(xfit,yfit,col='red')

}
```

bernoulli(0.2)



the delta method

- general results for the multivariate case: let $\boldsymbol{T} = (T_1, \ldots, T_k)$ denote a random vector with mean $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ and suppose we wish to approximate the variance of a differentiable function $g(\boldsymbol{T})$.

- first-order Taylor expansion:

$$
\begin{aligned}
g(\boldsymbol{t}) &\cong g(\boldsymbol{\theta}) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(t_i - \theta_i) \\
\mathbb{E}\big[g(\boldsymbol{T})\big] &\cong g(\boldsymbol{\theta}) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta})\mathbb{E}(T_i - \theta_i) \;=\; g(\boldsymbol{\theta}) \\
\text{var}\big[g(\boldsymbol{T})\big] &\cong \mathbb{E}\big[g(\boldsymbol{T}) - g(\boldsymbol{\theta})\big]^2 \;=\; \mathbb{E}\left[\sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(T_i - \theta_i)\right]^2 \\
&= \sum_{i=1}^{k} \big[g_i'(\boldsymbol{\theta})\big]^2 \text{var}(T_i) + 2 \sum_{1 \le i \ne j \ne k} g_i'(\boldsymbol{\theta}) g_j'(\boldsymbol{\theta}) \text{cov}(T_i, T_j)
\end{aligned}
$$

- theorem (**multivariate delta method**): suppose that Y_n is n-dimensional and

  $$\sqrt{n}\left(Y_n - \theta\right) \xrightarrow{d} N(0, \Sigma)$$

  then

  $$\sqrt{n}\left(g(Y_n) - g(\theta)\right) \xrightarrow{d} N\left(0, G(\theta_0)\Sigma G(\theta_0)'\right)$$

  where $G = \frac{\partial g(\theta)}{\partial \theta'}$.

Contents

Reference:

- Casella and Berger, Ch. 5

Exercises:

- 5.1–5.3, 5.5, 5.6, 5.8, 5.10, 5.13, 5.15, 5.22, 5.23, 5.25, 5.30, 5.31, 5.34, 5.36, 5.42.