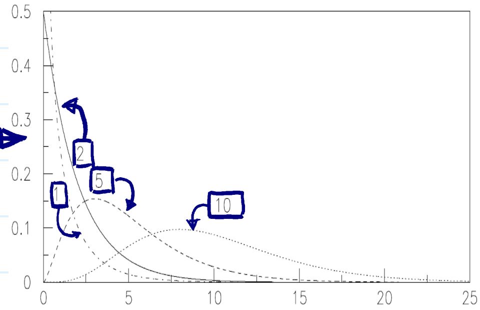
Definition 4.14 (Chi-square distribution χ^2_n, TBp.177)

- pdf: $f(x) = \begin{cases} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & x \leq 0 \end{cases}$ shape
- mgf: $(\frac{1}{1-2t})^{\frac{n}{2}}$
- mean: n
- variance: $2n$ intuition
- parameter: $n = 1, 2, 3, \ldots$



Notes:

1. n is called *degree of freedom* (自由度)
2. $\chi^2_n = \Gamma(\frac{n}{2}, \frac{1}{2})$ ← LNp.2-73~74
3. Let $X_1, \ldots, X_k$ be independent and $X_i \sim \chi^2_{n_i}$ ($= \Gamma(\frac{n_i}{2}, \frac{1}{2})$), then $Y = X_1 + \ldots + X_k \sim \chi^2_{n_1+\cdots+n_k}$ ← intuition, prove using mgf (Ec)

connection with normal

4. Let $Z \sim N(0,1)$, then $X = Z^2 \sim \chi^2_1$. ← Obtain the pdf of X from Z (Ec)
5. By 3 and 4, let $Z_1, \ldots, Z_n$ be i.i.d. $\sim N(0,1)$, then $Y = Z_1^2 + \cdots + Z_n^2 \sim \chi^2_n$. $Z_1^2, \ldots, Z_n^2 \overset{\text{i.i.d.}}{\sim} \chi^2_1$

Definition 4.15 (t distribution t_n, TBp.178)

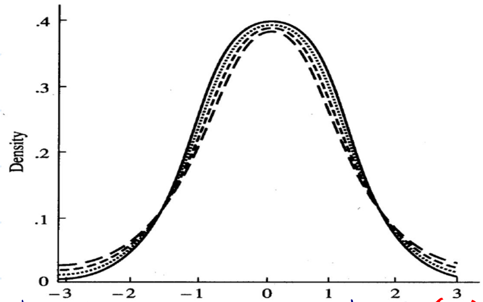
- pdf: $f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}.$ shape
- mgf: not exist, except at $t = 0$
- mean: 0, $(n > 1)$
- variance: $\frac{n}{n-2}$, $(n > 2)$ cf. → Cauchy(0,1) = t_1
- moments:

$$E(X^k) = \begin{cases} \frac{\Gamma(\frac{k+1}{2})\Gamma(\frac{n-k}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} n^{\frac{k}{2}}, & k < n \text{ and even} \\ 0, & k < n \text{ and odd} \end{cases}$$

← by Note 1 & LNp.74, Note 7 (Ec)

- parameter: $n = 1, 2, 3, \ldots$ (自由度)

t_5 (long dash), t_{10} (short dash), t_{30} (dot), $N(0,1)$ (solid)



For even k,

$EX^k = E\left(\frac{Z}{\sqrt{U/n}}\right)^k = E\left[(Z^2)^{\frac{k}{2}}\right] E\left(U^{-\frac{k}{2}}\right) \cdot n^{\frac{k}{2}}$

$\Gamma(\frac{1}{2}, \frac{1}{2}) = \chi^2_1$ $\quad \chi^2_m = \Gamma(\frac{m}{2}, \frac{1}{2})$

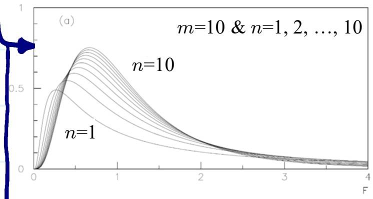
Notes:

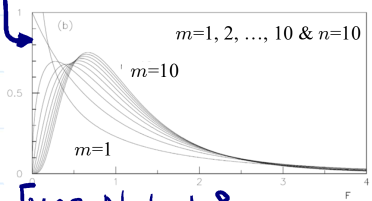
connection with normal

1. Let $Z \sim N(0,1)$ and $U \sim \chi^2_n$ be independent, then $\frac{Z}{\sqrt{U/n}} \sim t_n$. (degree of freedom)
   ① identify the pdf of $W = \sqrt{\frac{U}{n}}$ ② find the pdf of $\frac{Z}{W}$ (LNp.33) (Ec)
2. $f(x) = f(-x)$, i.e., t_n distribution is symmetric about zero
3. as $n \to \infty$, t_n tends to $N(0,1)$. (by LLN, Chapter 5)
4. t_n has heavier tail than $N(0,1)$

Definition 4.16 (F distribution $F_{m,n}$, TBp.179)

- pdf: $f(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{-\frac{m}{2}} x^{\frac{m}{2}-1} \left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ shape
- mgf: not exist, except at $t=0$
- mean: $\frac{n}{n-2}$, $(n>2)$ ← irrelevant to m
- variance: $\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, $(n>4)$
- moments: $E(X^k) = \frac{\Gamma(\frac{m+2k}{2})\Gamma(\frac{n-2k}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{n}{m}\right)^k$, $k < \frac{n}{2}$
- parameter: $m, n = 1, 2, 3, \ldots$ (自由度)





$E(X^k) = E\left[\left(\frac{U/m}{V/n}\right)^k\right] = E(U^k)E(V^{-k})\cdot\left(\frac{n}{m}\right)^k$

$\Gamma(\frac{m}{2},\frac{1}{2}) = \chi^2_m$ $\quad \chi^2_n = \Gamma(\frac{n}{2},\frac{1}{2})$

use Note 1 & LNp.74, Note 7 (Ec)

Notes:

1. Let $U \sim \chi^2_m$ and $V \sim \chi^2_n$ be independent, then $\frac{U/m}{V/n} \sim F_{m,n}$.

connection with normal ① find the pdfs of $Z = \frac{U}{m}$ & $W = \frac{V}{n}$ ② find the pdf of $\frac{Z}{W}$ (LNp.33) (Ec)

2. Let $X \sim t_n$, then $Y = X^2 \sim F_{1,n}$. → $Z \sim N(0,1)$, $U \sim \chi^2_n$) indep. $X = \frac{Z}{\sqrt{U/n}} \sim t_n \Rightarrow \frac{Z^2/1}{U/n} \sim F_{1,n}$ ($Z^2 \sim \chi^2_1$)

3. $X \sim F_{m,n} \Rightarrow X^{-1} \sim F_{n,m}$

Theorem 4.1 (distributions of sample mean and sample variance of i.i.d. normal, sec. 6.3)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\sim N(\mu, \sigma^2)$. Define ← unknown

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{(called } \textit{sample mean}\text{)}$$

$\boxed{E(\overline{X}_n) = \mu}$ (Ec) $\quad$ $Var(\overline{X}_n) = \frac{\sigma^2}{n} \xrightarrow{n\to\infty} 0$

r.v. used to understand μ $\quad$ cf. $E(X_1) = \mu$

transformation of $X_1, \cdots, X_n$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \quad \text{(called } \textit{sample variance}\text{)}$$

$-\mu+\mu$ $\quad$ $\boxed{E(S_n^2) = \sigma^2}$ (Ec)

$Y_i \sim N(0, \frac{n-1}{n}\sigma^2)$ (Ec) $\quad$ r.v. used to understand σ^2

cf. definition of variance, $Var(X_1) = E[(X_1 - E(X_1))^2] = \sigma^2$

Then,

1. $\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ and $\sqrt{n}(\overline{X}_n - \mu)/\sigma \sim N(0,1)$. ← by LNp.76, Notes 3 & 5.
   $Var(\overline{X}_n) \to 0$, $n \to \infty$, $\overline{X}_n \approx \mu$.
2. (TBp.195) The random variable $\overline{X}_n$ and the random vector $(X_1 - \overline{X}_n, X_2 - \overline{X}_n, \ldots, X_n - \overline{X}_n)$ are independent. ← $\overline{X}_n$ & $(Y_1, \cdots, Y_n)$ indep.
   prove in later slides
3. (TBp.196) $\overline{X}_n$ and S_n^2 are independently. $= \frac{1}{n-1}\sum_{i=1}^n Y_i^2$
4. (TBp.197) The distribution of $(n-1)S_n^2/\sigma^2$ is the chi-square distribution with $n-1$ degrees of freedom. → $Var(S_n^2) = Var\left(\frac{(n-1)S_n^2}{\sigma^2} \times \frac{\sigma^2}{n-1}\right) = \frac{\sigma^4}{(n-1)^2} Var\left(\frac{(n-1)S_n^2}{\sigma^2}\right) = \frac{\sigma^4}{(n-1)^2} \cdot 2(n-1) = \frac{2\sigma^4}{n-1} \xrightarrow{n\to\infty} 0$
   $(n-1)S_n^2 = \sum_{i=1}^n Y_i^2$, but $\sum Y_i = 0$
5. (TBp.198) $\frac{\sqrt{n}(\overline{X}_n - \mu)/\sigma}{\sqrt{(n-1)S_n^2/\sigma^2(n-1)}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}$. ← by LNp.78, Note 1
   cf. standardization

Proof of 2. The joint mgf of $\overline{X}_n$ and $(\underbrace{X_1-\overline{X}_n}_{Y_1}, \underbrace{X_2-\overline{X}_n}_{Y_2}, \ldots, \underbrace{X_n-\overline{X}_n}_{Y_n})$ is

$\overset{?}{=} M_{\overline{X}_n}(s) \times M_{Y_1,\ldots,Y_n}(t_1,\ldots,t_n)$

$$M(s,t_1,t_2,\ldots,t_n) = E\left\{e^{\left[s\overline{X}_n+\sum_{i=1}^n t_i(X_i-\overline{X}_n)\right]}\right\} = E\left\{e^{\left[\sum_{i=1}^n \underbrace{\left(\frac{s}{n}+t_i-\bar{t}\right)}_{\parallel\, a_i}X_i\right]}\right\}.$$

Let $a_i = \frac{s}{n}+t_i-\bar{t}$, $i=1,2,\ldots,n$. Then

Note: $X_1,\cdots,X_n \overset{\text{i.i.d}}{\sim} N(\mu,\sigma^2)$ $\Rightarrow a_1X_1,\cdots,a_nX_n \overset{\text{indep.}}{\sim} N(a_i\mu, a_i^2\sigma^2)$

$$\sum_{i=1}^n a_i = s, \quad \sum_{i=1}^n a_i^2 = \frac{s^2}{n}+\sum_{i=1}^n (t_i-\bar{t})^2.$$

Now we have

$$M(s,t_1,t_2,\ldots,t_n) = \prod_{i=1}^n M_{X_i}(a_i) = \prod_{i=1}^n \exp\left(\mu a_i + \frac{\sigma^2}{2}a_i^2\right)$$

$$= \exp\left(\mu\sum_{i=1}^n a_i + \frac{\sigma^2}{2}\sum_{i=1}^n a_i^2\right) = \exp\left[\mu s + \frac{\sigma^2}{2}\frac{s^2}{n} + \frac{\sigma^2}{2}\sum_{i=1}^n (t_i-\bar{t})^2\right]$$

$$= \underbrace{\exp\left(\mu s + \frac{\sigma^2}{2n}s^2\right)}_{\text{a function of } s \text{ only}} \underbrace{\exp\left[\frac{\sigma^2}{2}\sum_{i=1}^n (t_i-\bar{t})^2\right]}_{\text{a function of } (t_1,\cdots,t_n) \text{ only.}}$$

Thus, the joint mgf factorizes into product of the mgf of $\overline{X}_n$ and the mgf of $(X_1-\overline{X}_n, X_2-\overline{X}_n, \ldots, X_n-\overline{X}_n)$. $M(0,t_1,\cdots,t_n)$ $M(s,0,\cdots,0)$

Question 4.4: Are $(\underbrace{X_1-\overline{X}_n}_{Y_1}, \ldots, \underbrace{X_n-\overline{X}_n}_{Y_n})$ independent? [Hint: $\sum_{i=1}^n (\underbrace{X_i-\overline{X}_n}_{Y_i}) = 0$] $\sum_{i=1}^n Y_i = 0$.

Proof of 4. First note that

LNp.80

$$\frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\mu)^2 = \sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2.$$

$N(0,1) \sim T_i$ $\quad T_1^2,\cdots,T_n^2 \overset{\text{i.i.d}}{\sim} \chi_1^2$

Also,

$$\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^n (X_i \overset{-\overline{X}_n+\overline{X}_n}{} -\mu)^2}_{W\sim\chi_n^2} = \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\overline{X}_n)^2}_{U=\frac{(n-1)S_n^2}{\sigma^2}} + \underbrace{\left(\frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma}\right)^2}_{V\sim\chi_1^2}.$$

$$\boxed{+\frac{2}{\sigma^2}\sum_{i=1}^n (X_i-\overline{X}_n)(\overline{X}_n-\mu) = 0}$$

Since V and U are independent, (Why?)

$$M_W(t) = M_U(t)M_V(t)$$

and then

$(n-1)S_n^2 = \sum_{i=1}^n Y_i^2$

$$M_U(t) = \frac{M_W(t)}{M_V(t)} = \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}} = (1-2t)^{-\frac{n-1}{2}},$$

which is the mgf of a χ_{n-1}^2 distribution. Thus $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

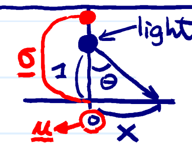
uniqueness Thm (LNp.46)

Question 4.5: Why degree of freedom $= n-1$, rather than n?

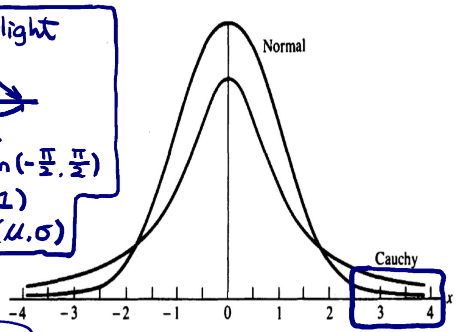
$(Y_1,\cdots,Y_n) \in \mathbb{R}^n$ but its possible values are on a $(n-1)$-dim subspace

Shape | Definition 4.17 (Cauchy distribution $C(\mu, \sigma)$, TBp.95)

- pdf: $f(x) = \frac{\sigma}{\pi}\frac{1}{\sigma^2+(x-\mu)^2}$, $x \in \mathbb{R}$.
- cdf: $\frac{1}{2} + \frac{1}{\pi}\tan^{-1}(\mu + \sigma x)$, $x \in \mathbb{R}$.
- mgf: not exist, except at $t = 0$
- chf: $e^{i\mu t - \sigma|t|}$
- mean: not exist
- variance: not exist
- parameter: $\mu \in \mathbb{R}$, $\sigma > 0$ ← cf.



$\theta \sim \text{Uniform}(-\frac{\pi}{2}, \frac{\pi}{2})$
$X \sim C(0,1)$
$\mu + \sigma X \sim C(\mu, \sigma)$



∵ tail is too heavy

Notes: location parameter (median) ; scale parameter

useful for modeling data with extreme values, e.g., black swan events in risk analysis

1. a heavy tail distribution
2. $C(0, 1) = t_1$ ← $X/\sqrt{Y^2} = X/|Y|$
3. Let X, Y be i.i.d. $\sim N(0, 1)$. Then, $X/Y \sim C(0, 1)$. ← LNp.33 (Ec)
4. $X \sim C(\mu, \sigma) \Rightarrow$ for $a, b \in \mathbb{R}$, $aX + b \sim C(a\mu + b, |a|\sigma)$ ← use chf (Ec)
5. Let $X_1, \ldots, X_k$ be independent, and $X_i \sim C(\mu_i, \sigma_i)$. Then $Y = X_1 + \cdots + X_k \sim C(\sum_{i=1}^{k} \mu_i, \sum_{i=1}^{k} \sigma_i)$ ← use chf (Ec)
6. by 4 and 5, let $X_1, \ldots, X_k$ be i.i.d $\sim C(\mu, \sigma)$, then $\overline{X}_k \sim C(\mu, \sigma)$. ← cf. → $Var(\overline{X}_k) = \sigma^2/k$, when σ^2 exists

Some other distributions

Log-normal (TBp. 69)
Weibull (TBp. 69)
Double exponential (TBp. 111)
Logistic
Pareto (TBp. 323)
Maxwell (TBp. 121)

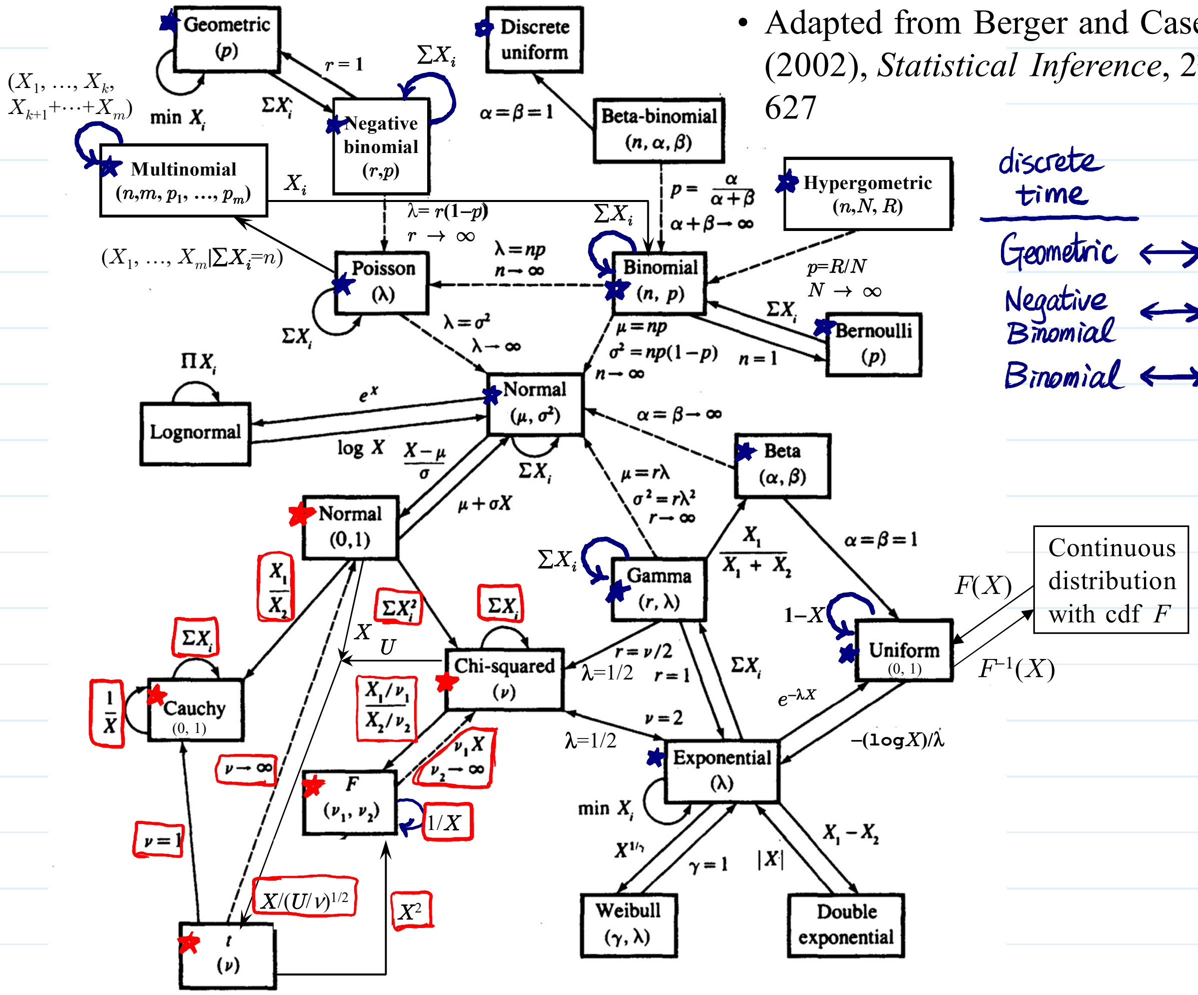
In this chapter, you should learn

1. random phenomenon behind each distribution
2. statistical modeling (assigning a distribution) of data
3. relationship among distributions
4. meaning of parameters in each distribution
5. HOW to derive cdf/mgf/chf/mean/variance from pdf/pmf (optional) ← Ec's

but you are not necessary to

1. memorize their pdf/pmf/cdf/mgf/chf/mean/variance/…

❖ **Reading**: textbook, 2.1.1~2.1.5, 2.2.1~2.2.4, chapter 6
Roussas, 3.2, 3.3, 3.4, 5.2, 6.3, 7.3



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).

- Adapted from Berger and Casella (2002), *Statistical Inference*, 2nd Ed., p.2-627

discrete time		continuous time
Geometric	$\leftrightarrow$	Exponential
Negative Binomial	$\leftrightarrow$	Gamma
Binomial	$\leftrightarrow$	Poisson

Chapter 5

Outline

- 3 types of convergence
  - a.s., in prob., in dist.
- law of large number 大數法則
- central limit theorem 中央極限定理.

value	prob.
0/3	1/8
1/3	3/8
2/3	3/8
3/3	1/8

compare the degrees of uncertainty

Question 5.1

1. repeat flipping a fair coin **2 or 3 times**. Can you accurately predict the average appearance of heads?
   c.f.
2. repeat flipping a fair coin **many many times**. What will you predict the average appearance of heads? say, 10000 times

Note. 規律 average ≈ 0.5 ← interpretation of mean (LNp.41) 隨機

1. Some *deterministic* patterns emerge from *random* phenomena when more and more data are collected, i.e., more and more information is gathered. e.g., $n = 3$ in 1, $n = 10000$ in 2 c.f.
2. In the following, $n \to \infty$ can be interpreted as *sample size of data is large enough.* ⇒ "asymptotic" (大樣本)

- three types of convergence

$$\lim_{n\to\infty} a_n = a \Leftrightarrow \lim_{n\to\infty} |a_n - a| = 0$$

Definition 5.1 (converge almost surely, TBp. 178)

A sequence of random variables $\{Z_n : \Omega \to \mathbb{R}\}$ is said to **converge almost surely** to a random variable $Z : \Omega \to \mathbb{R}$, and denoted as $Z_n \xrightarrow{\text{a.s.}} Z$, if for any $\epsilon > 0$, $\boxed{Z_n \xleftrightarrow{\text{alike}} Z \text{ as } n\to\infty}$

$$P\left(\left\{\omega \in \Omega : \lim_{n\longrightarrow\infty} |Z_n(\omega) - Z(\omega)| < \epsilon\right\}\right) = 1.$$

$\boxed{\text{converge pointwise}}$

A: $P(A)=1$, $P(A^c)=0$

$\boxed{\lim_{n\to\infty} |Z_n(\omega)-Z(\omega)| = 0 \Rightarrow \lim_{n\to\infty} Z_n(\omega)=Z(\omega)}$

3/5

Definition 5.2 (converge in probability, TBp. 178)

A sequence of random variables $\{Z_n : \Omega \to \mathbb{R}\}$ is said to **converge in probability** to a random variable $Z : \Omega \to \mathbb{R}$, and denoted as $Z_n \xrightarrow{P} Z$, if for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega : |Z_n(\omega) - Z(\omega)| < \epsilon\}\right) = 1.$$

A_n

$P(\cdot)$: $A_1, A_2, \ldots, A_n, \ldots \longrightarrow 1$

$\ni \omega$, $\not\ni \omega$, …, $\not\ni \omega$, …, $\ni \omega$, …, $\not\ni \omega$, $\ni \omega$, …

cont. r.v., $P(\{\omega\})=0$

cf.

$\boxed{Z=\alpha}$

$Z_n \xrightarrow{a.s.} \alpha$, α : a constant. $\quad Z_n \xrightarrow{a.s.} Z \Leftrightarrow Y_n = Z_n - Z \xrightarrow{a.s.} 0$

$Z_n \xrightarrow{P} \alpha$, α : a constant. $\quad Z_n \xrightarrow{P} Z \Leftrightarrow Y_n = Z_n - Z \xrightarrow{P} 0$