








# Statistica piețelor financiare

## Seminarul 2: Distribuții clasice și fapte stilizate



Daniel Traian PELE

Academia de Studii Economice din București
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Academia Română, Institutul de Prognoză Economică
MSCA Digital Finance

# Acronime folosite în acest material

- **ACF**: Autocorrelation Function (funcția de autocorelație)
- **AI**: Artificial Intelligence (inteligență artificială)
- **AIC**: Akaike Information Criterion (criteriul informațional Akaike)
- **BET**: Bucharest Exchange Trading (index) — indicele de referință al Bursei de Valori București
- **CDF**: Cumulative Distribution Function (funcția de repartiție)
- **CLT**: Central Limit Theorem (teorema limită centrală)
- **EODHD**: EOD Historical Data (market-data provider) — furnizor de date de piață
- **JB**: Jarque–Bera (test) — testul de normalitate Jarque–Bera
- **MLE**: Maximum Likelihood Estimation (estimarea prin metoda verosimilității maxime)
- **OMV**: Österreichische Mineralölverwaltung (Administrația austriacă a uleiurilor minerale, grupul OMV)
- **QQ**: Quantile–Quantile (plot) — graficul cuantilă–cuantilă
- **SE**: Standard Error (eroare standard)

# Întrebarea de azi și structura seminarului

- Întrebarea seminarului
  - **cît de departe sînt randamentele zilnice reale de distribuția Normală și cum măsurăm acest lucru?**
- Exercițiile: în ordinea din Harta exercițiilor (slide-ul următor)
- Rezultate de obținut
  - răspunsurile la exercițiile A; tabelul sau figura cerute la fiecare exercițiu B
  - un răspuns scurt și argumentat la C1; erorile găsite în răspunsul AI din C2 și răspunsul corectat
  - frazele de interpretare cerute la fiecare exercițiu
- Seminarul are rol de exercițiu și nu se notează; rezolvările exercițiilor [Propus] se discută la seminar
- Notebook-ul de azi: deschideți notebook-ul seminarului în Google Colab; fiecare cerință indică secțiunea
- Contact: prof. Daniel Traian Pele, danpele@ase.ro

# Harta exercițiilor

| Cerința | Întrebarea | Tipul | Model |
|---|---|---|---|
| A1, A2 | cît de probabilă este o pierdere zilnică mare în ipoteza distribuției Normale? | Rezolvat, Propus | A1 |
| A3, A4 | care este prețul după un an și după zece ani, dacă prețurile sînt lognormale? | Rezolvat, Propus | A3 |
| A5, A6 | cînd este o sumă aproximativ Normală (CLT)? | Rezolvat, Propus | A5 |
| A7, A8 | statistica Jarque–Bera și momentele unei distribuții Student-$t$ | Rezolvat, Propus | A7 |
| B1, B2 | cît de departe de normalitate sînt randamentele zilnice și cît de precis este estimată boltirea? | Rezolvat, Propus | B1 |
| B3, B4 | se potrivește mai bine o distribuție Student-$t$? cîte zile de 4 sigma? | Rezolvat, Propus | B3 |
| B5–B7 | gaussianitate agregată, volatility clustering, efect de levier și asimetrie | Rezolvat, Propus | B5 |
| C1, C2 | devin mai subțiri cozile Bitcoin? ce este greșit într-un răspuns generat de AI? | Propus | B3, B5 |

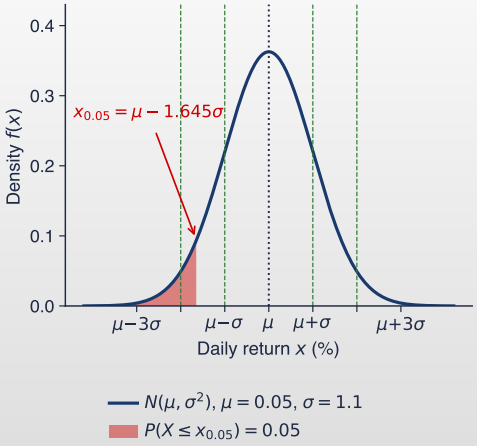
- **[Rezolvat]**: rezolvarea completă în slide-uri și în notebook, un model de urmat; **[Propus]**: îl rezolvați dumneavoastră, după model
- **A**: calcule pe hîrtie; **B**: date reale, în notebook; **C**: întrebări deschise și critica unui răspuns generat de AI

# Datele folosite

| Seria | Sursa | Prețul folosit | Perioada |
|---|---|---|---|
| S&P 500, DAX, BET | EODHD | close | 2010–2026 |
| Bitcoin | EODHD | close, 7 zile pe săptămînă | 2014–2026 |
| OMV Petrom (SNP) | EODHD | adjusted close | 2010–2026 |

- Randamente logaritmice zilnice în %: $r_t = 100\,(\ln P_t - \ln P_{t-1})$
- Weekendurile și prețurile de închidere repetate din zilele de sărbătoare se elimină (cu excepția Bitcoin); fiecare serie are calendarul ei
- În notebook: `returns('sp500')`, `moments(r)`, `t_fit(r)`; nu este nevoie de cont sau de cheie

# Noțiuni necesare azi (1/15): distribuția Normală




- $X \sim N(\mu, \sigma^2)$ are densitatea

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)$$

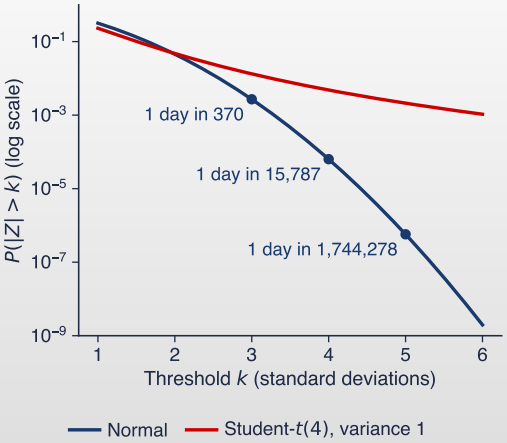
- $\mu$: media (centrul clopotului); $\sigma^2$: varianța (lățimea lui); $\exp(x) = e^x$
  - aria de sub $f$ dintre două puncte este o probabilitate
- Standardizarea: distanța față de medie, măsurată în abateri standard

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1), \qquad P(X \leq x) = \Phi\Big(\frac{x-\mu}{\sigma}\Big)$$

- $\Phi$: funcția de repartiție (CDF, cumulative distribution function) a lui $N(0,1)$
  - cuantila de ordin $\alpha$: $x_\alpha = \mu + \sigma z_\alpha$, cu $z_\alpha = \Phi^{-1}(\alpha)$ și $P(X \leq x_\alpha) = \alpha$
  - $z_{0,05} = -1{,}645$, $z_{0,025} = -1{,}96$, $z_{0,01} = -2{,}326$; aria roșie din grafic: $\alpha = 0{,}05$

SFM_ch2_seminar

# Noțiuni necesare azi (2/15): zile de $k$ sigma




- O **zi de $k$ sigma**: un randament aflat la mai mult de $k$ abateri standard de medie

$$|r_t - \bar{r}| > k\, s$$

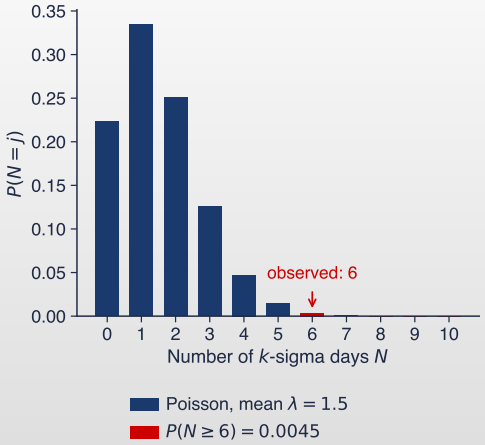
- $r_t$: randamentul logaritmic din ziua $t$; $\bar{r}$, $s$: media și abaterea standard de selecție
- Conform distribuției Normale: probabilitatea și **numărul așteptat** în $n$ zile

$$P(|Z| > k) = 2\big(1 - \Phi(k)\big), \qquad E[N] = n\, P(|Z| > k)$$

- $N$: numărul zilelor de $k$ sigma dintre cele $n$ zile
  - $P(|Z| > 4) = 6{,}3 \times 10^{-5}$: aproximativ o zi din 15 787, o dată la 63 de ani de tranzacționare (252 de zile pe an)
  - o distribuție $t(4)$ cu aceeași variață face zilele de 4 sigma de circa 30 de ori mai frecvente

SFM_ch2_seminar

# Noțiuni necesare azi (3/15): numărul Poisson și p-value-ul lui




- Evenimente rare, fiecare zi fiind o încercare independentă: $N$ urmează aproximativ o distribuție **Poisson**

$$P(N = j) = e^{-\lambda} \frac{\lambda^j}{j!}, \qquad j = 0, 1, 2, \ldots$$

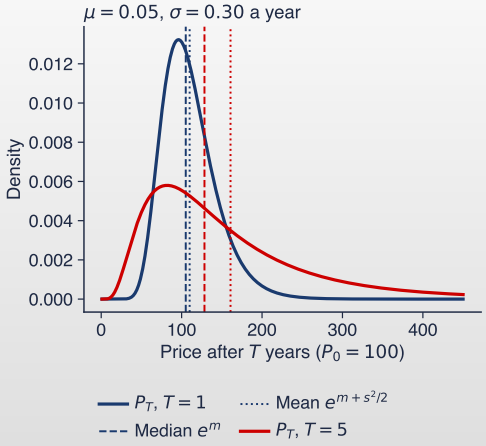
- $\lambda = n\,P(|Z| > k)$: numărul așteptat conform modelului Normal
  - $j! = 1 \cdot 2 \cdots j$; media și varianța lui $N$ sînt ambele egale cu $\lambda$
- **p-value-ul Poisson** al unui număr observat $N_{\mathbf{obs}}$

$$p = P(N \geq N_{\mathbf{obs}}) = 1 - \sum_{j=0}^{N_{\mathbf{obs}}-1} e^{-\lambda} \frac{\lambda^j}{j!}$$

- probabilitatea de a observa cel puțin numărul observat, dacă modelul Normal ar fi corect
  - un $p$ foarte mic: numărul se explică prin model, nu prin noroc; în grafic, $\lambda = 1{,}5$ și 6 zile observate dau $p = 0{,}0045$

SFM_ch2_seminar

# Noțiuni necesare azi (4/15): prețuri lognormale




- $Y$ este **lognormală**, $Y \sim LN(m, s^2)$, dacă $\ln Y \sim N(m, s^2)$

$$\text{mediana} = e^m, \qquad E[Y] = e^{m+s^2/2}$$

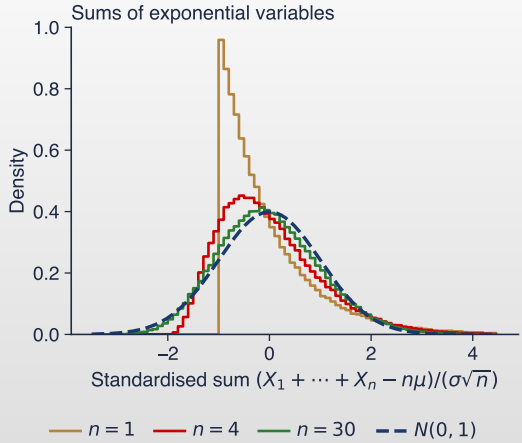
- $m$, $s^2$: media și varianța lui $\ln Y$, nu ale lui $Y$
  - $P(Y \le y) = \Phi\big((\ln y - m)/s\big)$, pentru că $Y \le y$ exact atunci cînd $\ln Y \le \ln y$
  - media depășește mediana: coada dreaptă lungă o trage în sus
- Randamente logaritmice anuale i.i.d. $N(\mu, \sigma^2)$; prețul după $T$ ani

$$P_T = P_0\, e^{r_1 + \cdots + r_T}, \qquad \ln \frac{P_T}{P_0} \sim N(T\mu,\ T\sigma^2)$$

- i.i.d.: independente și identic distribuite; $P_0$: prețul de azi; $T$: orizontul, în ani
  - deci $m = \ln P_0 + T\mu$ și $s = \sigma\sqrt{T}$: împrăștierea crește cu orizontul (curba roșie, $T = 5$)

SFM_ch2_seminar

# Noțiuni necesare azi (5/15): teorema limită centrală



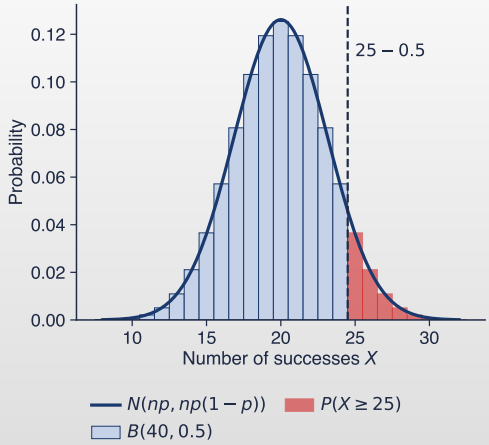

- **CLT** (central limit theorem, teorema limită centrală): pentru $X_1, \ldots, X_n$ i.i.d., cu media $\mu$ și varianța finită $\sigma^2$

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \approx N(0, 1)$$

- echivalent, suma este aproximativ $N(n\mu,\ n\sigma^2)$
  - aproximarea se îmbunătățește cînd $n$ crește
- Trei condiții
  - independență; distribuție identică; varianță finită
  - cu varianță infinită (de exemplu Student-$t$ cu $\nu \leq 2$), limita nu este Normală (Capitolul 3)
- În grafic: sume de variabile exponențiale, foarte asimetrice
  - pentru $n = 1$ forma este departe de clopot; cu $n = 30$, suma este deja apropiată de $N(0, 1)$

SFM_ch2_seminar

# Noțiuni necesare azi (6/15): distribuția binomială și aproximarea Normală




- $X \sim B(n,p)$: numărul de succese în $n$ încercări independente, cu probabilitatea $p$

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \qquad k=0,1,\ldots,n$$

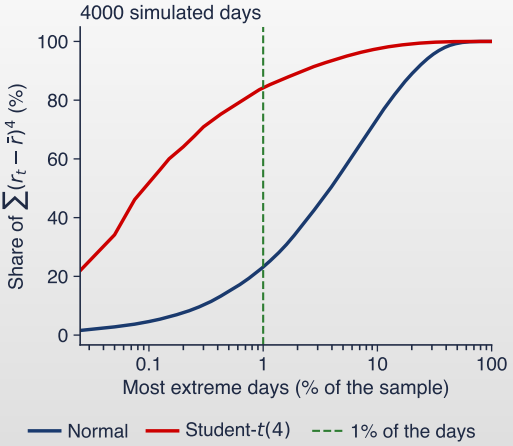
- $\binom{n}{k}=\frac{n!}{k!\,(n-k)!}$: numărul de moduri de a alege cele $k$ succese
  - media $E[X]=np$, varianța $\mathrm{Var}(X)=np(1-p)$
  - $X$ este o sumă de $n$ variabile 0/1 independente, deci se aplică CLT
- Aproximarea Normală cu **corecția de continuitate**

$$P(X \geq k) \approx 1-\Phi\Big(\frac{k-0{,}5-np}{\sqrt{np(1-p)}}\Big)$$

- întregul $k$ acoperă intervalul $[k-0{,}5;\, k+0{,}5]$ (linia întreruptă din grafic)
  - bună cînd $np(1-p)$ este mare (regulă practică: peste 10)

SFM_ch2_seminar

# Noțiuni necesare azi (7/15): asimetria și excesul de boltire




- Momentele centrate de selecție și cele două măsuri ale formei

$$m_k = \frac{1}{n}\sum_{t=1}^{n}(r_t - \bar{r})^k, \qquad S = \frac{m_3}{m_2^{3/2}}, \qquad K = \frac{m_4}{m_2^2} - 3$$

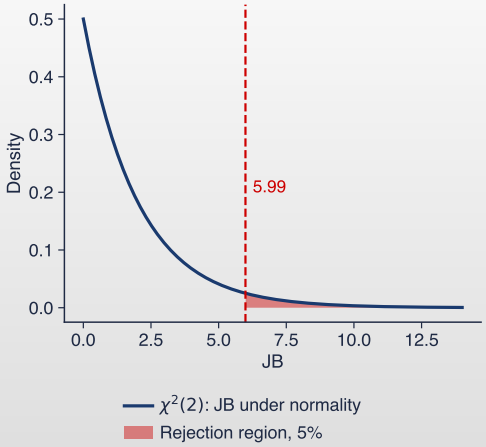
- $m_2$: varianța; $S < 0$: o coadă stîngă mai lungă; $K > 0$: cozi mai groase decît la distribuția Normală
  - pentru distribuția Normală $S = K = 0$; boltirea propriu-zisă este $m_4/m_2^2 = K + 3$
- Erorile standard în ipoteza de normalitate și statisticile $z$

$$\mathrm{SE}(S) = \sqrt{6/n}, \qquad \mathrm{SE}(K) = \sqrt{24/n}$$

- $z_S = S/\mathrm{SE}(S)$, $z_K = K/\mathrm{SE}(K)$: fiecare aproximativ $N(0,1)$ în ipoteza de normalitate
- În grafic: partea din $\sum(r_t - \bar{r})^4$ care provine din zilele cele mai extreme
  - cozi groase: 1% din zile dau cea mai mare parte din $m_4$, deci $K$ este fragil

SFM_ch2_seminar

# Noțiuni necesare azi (8/15): testul Jarque–Bera



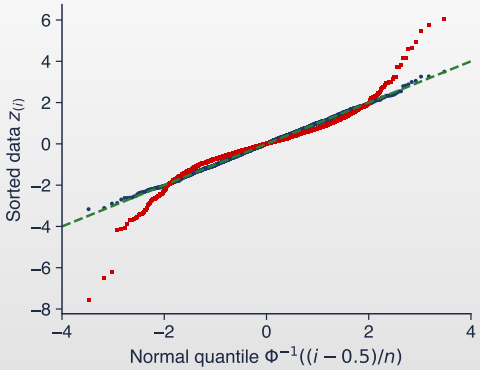

- **Jarque–Bera** Jarque & Bera (1987): o singură statistică pentru ambele măsuri ale formei

$$\text{JB} = \frac{n}{6}\left(S^2 + \frac{K^2}{4}\right) = z_S^2 + z_K^2$$

- zero pentru o formă Normală, mare cînd $S$ sau $K$ este departe de 0
  - $H_0$: datele provin dintr-o distribuție Normală
- Distribuția de referință: în ipoteza $H_0$, $\text{JB} \approx \chi^2(2)$
  - $\chi^2(2)$: distribuția hi-pătrat cu 2 grade de libertate, suma pătratelor a două variabile $N(0,1)$
  - respingem normalitatea la 5% dacă $\text{JB} > 5{,}99$ (regiunea roșie); p-value $= P(\chi^2(2) > \text{JB}) = e^{-\text{JB}/2}$
- Cu mii de zile, chiar și abateri mici dau un JB foarte mare

SFM_ch2_seminar

# Noțiuni necesare azi (9/15): QQ plot-ul



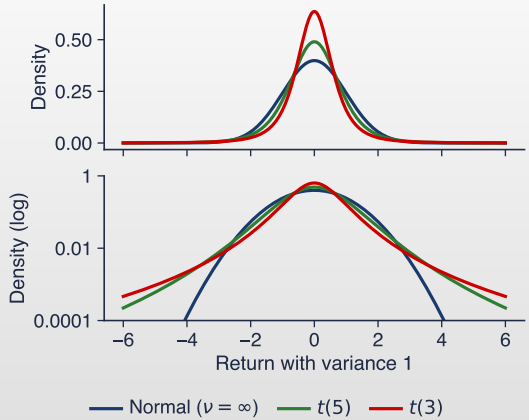

- **QQ plot**: datele ordonate față de cuantilele unui model

$$\left(F^{-1}\left(\frac{i-0{,}5}{n}\right),\ z_{(i)}\right), \qquad i = 1, \ldots, n$$

- $z_{(i)}$: al $i$-lea cel mai mic randament standardizat, $z_t = (r_t - \bar{r})/s$
  - $F^{-1}$: funcția cuantilă a modelului; pentru $N(0,1)$, $F^{-1} = \Phi^{-1}$
  - $(i - 0{,}5)/n$: proporția datelor aflate sub al $i$-lea punct
- Interpretarea
  - punctele pe dreapta de 45 de grade: modelul se potrivește (albastru)
  - o formă de S, jos la stînga și sus la dreapta: cozi mai groase decît ale modelului (roșu)

SFM_ch2_seminar

# Noțiuni necesare azi (10/15): distribuția Student-$t$




- **Student-$t(\nu)$** Student (1908): o densitate simetrică, cu cozi de tip putere

$$f(x) \propto \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

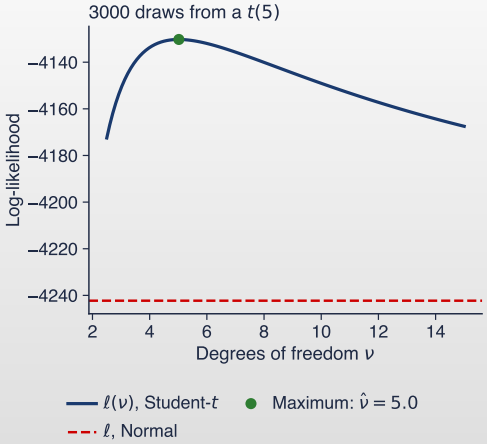
- $\nu > 0$: gradele de libertate; $\nu$ mai mic, cozi mai groase; $\nu \to \infty$: distribuția Normală
  - $\propto$: proporțional cu (o constantă face aria egală cu 1)
- Momentele lui $T \sim t(\nu)$; sub aceste praguri sînt infinite

$$\text{Var}(T) = \frac{\nu}{\nu - 2} \ (\nu > 2), \qquad K = \frac{6}{\nu - 4} \ (\nu > 4)$$

- Pentru randamente: $r = m + s\,T$, cu poziția $m$ și scala $s$
  - abaterea standard $s\sqrt{\nu/(\nu-2)}$; în grafic, fiecare $t$ are varianța 1
  - un centru mai înalt și cozi mai groase decît la distribuția Normală (scara logaritmică, jos)

SFM_ch2_seminar

# Noțiuni necesare azi (11/15): verosimilitatea maximă și AIC




- **MLE** (maximum likelihood estimation, estimarea prin verosimilitate maximă): parametrii care fac datele cele mai probabile

$$\ell(\theta) = \sum_{t=\mathbf{1}}^{n} \ln f(r_t; \theta), \qquad \hat{\theta} = \arg\max_{\theta} \ell(\theta)$$

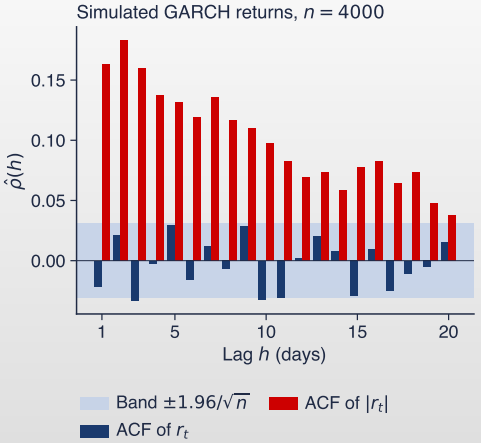
- $\theta$: parametrii, pentru $t$: $(\nu, m, s)$; $f$: densitatea modelului
  - $\ell$: log-verosimilitatea; notația $\hat{}$ indică o estimare, de exemplu $\hat{\nu}$
- **AIC** (Akaike information criterion, criteriul informațional Akaike): potrivirea penalizată cu numărul de parametri

$$\text{AIC} = 2k - 2\ell(\hat{\theta})$$

- $k$: numărul de parametri (2 pentru distribuția Normală, 3 pentru $t$); se preferă valoarea AIC mai mică
  - în grafic: $\ell(\nu)$ are maximul aproape de valoarea adevărată $\nu = 5$, mult peste $\ell$ al distribuției Normale

SFM_ch2_seminar

# Noțiuni necesare azi (12/15): ACF și testul Ljung–Box




- **ACF** (autocorrelation function, funcția de autocorelație): corelația randamentelor aflate la $h$ zile distanță

$$\hat{\rho}(h) = \frac{\sum_{t=h+1}^{n}(r_t - \bar{r})(r_{t-h} - \bar{r})}{\sum_{t=1}^{n}(r_t - \bar{r})^2}$$

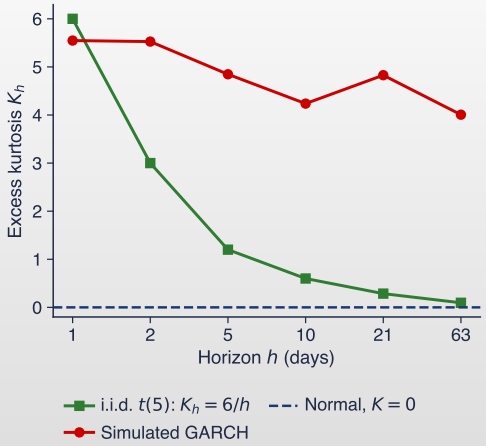
- $h$: lagul; banda de 95% în ipoteza de independență: $\pm 1{,}96/\sqrt{n}$
  - aceeași formulă cu $|r_t|$ în locul lui $r_t$ măsoară dependența mărimii variațiilor
- **Ljung–Box** Ljung & Box (1978): primele $m$ laguri testate împreună

$$Q(m) = n(n+2)\sum_{h=1}^{m}\frac{\hat{\rho}(h)^2}{n-h} \;\approx\; \chi^2(m) \text{ în ipoteza } H_0$$

- $H_0$: nicio autocorelație pînă la lagul $m$; respingem la 5% dacă $Q(10) > \chi^2_{0,95}(10) = 18{,}31$
  - în grafic: $r_t$ aproape necorelate, $|r_t|$ clar autocorelate: **volatility clustering**

SFM_ch2_seminar

# Noțiuni necesare azi (13/15): gaussianitatea agregată




- **Randamentele pe $h$ zile**: sume de $h$ randamente logaritmice zilnice, pe blocuri disjuncte

$$r_j^{(h)} = \sum_{t=(j-\mathbf{1})h+\mathbf{1}}^{jh} r_t, \qquad j = 1, \ldots, \lfloor n/h \rfloor$$

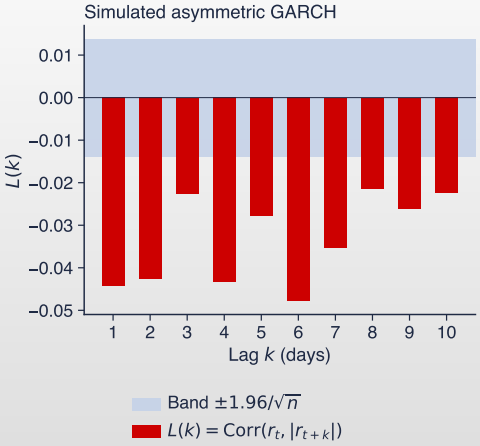
- $\lfloor n/h \rfloor$: numărul de blocuri complete; mai puține date cînd $h$ crește
- Pentru randamente zilnice i.i.d., măsurile formei scad cu $h$

$$S_h = \frac{S}{\sqrt{h}}, \qquad K_h = \frac{K}{h}$$

- **Gaussianitatea agregată**: randamentele pe orizonturi mai lungi sînt mai apropiate de distribuția Normală
  - volatility clustering încetinește scăderea (curba roșie din grafic)
  - cu puține blocuri, $\mathrm{SE}(K_h) = \sqrt{24/\lfloor n/h \rfloor}$ este mare, iar testele își pierd puterea

SFM_ch2_seminar

# Noțiuni necesare azi (14/15): efectul de levier și asimetria cîștig/pierdere




- **Efectul de levier**: o scădere azi este urmată de variații mai mari
$$L(k) = \mathrm{Corr}\big(r_t,\ |r_{t+k}|\big) < 0$$
- $k$: numărul de zile înainte; un $L(k)$ negativ, în afara benzii $\pm 1{,}96/\sqrt{n}$
  - pentru o firmă, o scădere mărește datoria în raport cu capitalul propriu, deci riscul
- **Asimetria cîștig/pierdere**: pierderile mari sînt mai ample și mai frecvente decît cîștigurile mari
  - comparăm $|q_{\mathbf{0{,}01}}|$ cu $q_{\mathbf{0{,}99}}$; $q_p$: cuantila de selecție de ordin $p$
- **Testul binomial**: dintre $N = D + U$ zile extreme, $D$ sînt scăderi și $U$ creșteri

  $H_{\mathbf{0}}:\ D \sim B(N; 0{,}5), \qquad p = P\big(|D - N/2| \geq |D_{\mathbf{obs}} - N/2|\big)$
- Un $p$ mic: scăderile și creșterile nu sînt la fel de probabile

SFM_ch2_seminar

# Noțiuni necesare azi (15/15): fapte stilizate, bootstrap și corelația rangurilor

- **Faptele stilizate** Cont (2001): proprietăți statistice comune majorității activelor și perioadelor
  - cozi groase; lipsa autocorelației liniare; volatility clustering; gaussianitate agregată; efect de levier; asimetrie cîștig/pierdere
- **Bootstrap**: precizia unei statistici $\hat{\theta}$ (de exemplu $S$ sau $K$), fără o formulă
  - extragem cu întoarcere $n$ zile din eșantion și recalculăm $\hat{\theta}_b^*$; repetăm pentru $b = 1, \ldots, B$
  - intervalul de 95%: percentilele 2,5% și 97,5% ale lui $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$; abaterea lor standard este SE bootstrap
  - spre deosebire de $\sqrt{24/n}$, nu presupune date Normale
- **Corelația rangurilor Spearman** pentru $n$ perechi $(x_i, y_i)$, de exemplu (anul, $\hat{\nu}$)

$$\hat{\rho}_S = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

- $d_i$: diferența dintre rangul lui $x_i$ și rangul lui $y_i$ (rangul 1 = cea mai mică valoare)
  - $\hat{\rho}_S \in [-1, 1]$; $+1$: o legătură monoton crescătoare; p-value-ul lui testează $H_0$: nicio legătură monotonă
  - cu puține puncte (cîțiva ani), doar un trend foarte puternic este semnificativ

# A1: o pierdere zilnică mare în ipoteza distribuției Normale [Rezolvat]

## Cerință

- Randamentele logaritmice zilnice ale unui indice sînt $N(\mu, \sigma^2)$, cu $\mu = 0{,}05\%$ și $\sigma = 1{,}1\%$; un an are 252 de zile de tranzacționare.
- 1. Calculați probabilitatea unei pierderi zilnice mai mari de 2% și numărul așteptat de asemenea zile pe an.
- 2. Calculați cuantila de 1% a randamentului zilnic.
- 3. Calculați numărul așteptat de zile de 4 sigma în 10 ani.
- Raportați: o probabilitate, un număr de zile, o cuantilă și un număr așteptat.

## Rezolvare

- 1. $z = (-2 - 0{,}05)/1{,}1 = -1{,}86$;
  $P = \Phi(-1{,}86) = 3{,}12\%$; $252 \times 3{,}12\% = 7{,}9$ zile pe an
- 2. $x_{\mathbf{0{,}01}} = 0{,}05 + 1{,}1 \times (-2{,}326) = -2{,}51\%$
- 3. $P(|Z| > 4) = 6{,}3 \times 10^{-\mathbf{5}}$; în 2 520 de zile: 0,16 zile, deci majoritatea deceniilor n-ar avea niciuna
- Cursul 2 arată că indicii reali au o asemenea zi la cîteva luni.

înapoi Lista exercițiilor

# A2: aceleași întrebări pentru Bitcoin [Propus]

Rezolvarea se discută la seminar

## Cerință

- Randamentele logaritmice zilnice ale Bitcoin sînt $N(0{,}12\%, 3{,}6\%^2)$; Bitcoin se tranzacționează 365 de zile pe an. Model: A1.
- 1. Calculați probabilitatea unei pierderi zilnice mai mari de 10% și numărul așteptat de asemenea zile pe an.
- 2. Calculați cuantila de 1% a randamentului zilnic.
- 3. Calculați numărul așteptat de zile de 4 sigma în 10 ani.
- Raportați: o probabilitate, un număr de zile, o cuantilă și un număr așteptat.

Înapoi Lista exercițiilor

# A3: prețuri lognormale după un an și după zece ani [Rezolvat]

## Cerință

- O acțiune costă 100 de lei; randamentele ei logaritmice anuale sînt i.i.d. $N(0{,}07, 0{,}18^2)$.
- 1. Calculați mediana și media prețului după un an.
- 2. Calculați probabilitatea ca prețul după un an să fie sub 100 de lei.
- 3. Calculați cuantila de 5% a prețului după un an.
- 4. Repetați 1–3 pentru zece ani.
- Raportați: opt valori și o frază despre orizont.

## Rezolvare

- 1. Mediana $100e^{\mathbf{0{,}07}} = 107{,}25$; media $100e^{\mathbf{0{,}07}+\mathbf{0{,}0162}} = 109{,}00$
- 2. $P = \Phi(-0{,}07/0{,}18) = \Phi(-0{,}389) = 34{,}9\%$
- 3. $100e^{\mathbf{0{,}07}+\mathbf{0{,}18}\times(\mathbf{-1{,}645})} = 79{,}77$
- 4. $T = 10$: $10\mu = 0{,}7$, $\sqrt{10}\sigma = 0{,}569$; mediana 201,38, media 236,79, $P(\text{pierdere}) = 10{,}9\%$, cuantila de 5% 78,96
- O pierdere devine mai puțin probabilă în timp, dar cuantila de 5% nu se îmbunătățește: 79,77 după un an, 78,96 după zece ani.

◂ Înapoi ▸ Lista exercițiilor

# A4: derivarea mediei $e^{m+s^2/2}$ [Propus]

Rezolvarea se discută la seminar

## Cerință

- Fie $X \sim N(m, s^2)$ și $Y = e^X$. Model: A3; exercițiu de tipul celor din Borak, Härdle & López-Cabrera (2013).
- 1. Scrieți $E[e^X] = \int e^x f(x)\,dx$ și formați pătratul perfect în exponent pentru a arăta că $E[e^X] = e^{m+s^2/2}$.
- 2. Folosiți aceeași idee pentru $E[e^{2X}]$ și deduceți $\mathrm{Var}(Y) = (e^{s^2} - 1)e^{2m+s^2}$.
- 3. Pentru $m = 0{,}07$, $s = 0{,}18$, calculați media și mediana lui $Y$ și diferența lor.
- Raportați: cele două derivări și o frază care leagă diferența de volatility drag din Capitolul 1.

◂ Înapoi ▸ Lista exercițiilor

# A5: zilele de creștere dintr-un an (aproximarea Normală) [Rezolvat]

## Cerință

- Un indice crește într-o zi dată cu probabilitatea $p = 0{,}53$, independent de la o zi la alta; un an are 252 de zile de tranzacționare.
- 1. Precizați distribuția numărului $X$ de zile de creștere, media și abaterea standard a lui $X$.
- 2. Aproximați $P(X \geq 140)$ cu distribuția Normală și corecția de continuitate.
- 3. Comparați cu valoarea binomială exactă din notebook.
- Raportați: două momente, două probabilități, o frază despre calitatea aproximării.

## Rezolvare

- 1. $X \sim B(252, 0{,}53)$; $E[X] = 133{,}56$, ab. std. $= \sqrt{252 \times 0{,}53 \times 0{,}47} = 7{,}92$
- 2. $z = (139{,}5 - 133{,}56)/7{,}92 = 0{,}750$; $P \approx 1 - \Phi(0{,}750) = 22{,}67\%$
- 3. Exact: 22,69%
- $np(1-p) \approx 63$ este mare, deci aproximarea dată de CLT este foarte bună aici.

◂ Înapoi ▸ Lista exercițiilor

# A6: este Normal un randament lunar? [Propus]

Rezolvarea se discută la seminar

## Cerință

- Randamentele logaritmice zilnice sînt i.i.d., cu media 0,04% și abaterea standard 1,2%. Model: A5.
- 1. Precizați distribuția aproximativă a randamentului logaritmic pe 21 de zile și cei doi parametri ai ei.
- 2. Calculați probabilitatea unei pierderi pe 21 de zile mai mari de 10% și a unei pierderi pe 5 zile mai mari de 10%.
- 3. Stabiliți dacă CLT se mai aplică atunci cînd randamentele zilnice sînt Student-$t$ cu $\nu = 3$.
- 4. Stabiliți dacă CLT se aplică pentru $\nu = 1{,}5$.
- Raportați: două distribuții, două probabilități, două răspunsuri, fiecare cu o justificare.

Înapoi Lista exercițiilor

# A7: Jarque–Bera pe hîrtie [Rezolvat]

## Cerință

- O serie de $n = 4000$ de randamente zilnice are asimetria $S = -0{,}6$ și excesul de boltire $K = 12$.
- 1. Calculați erorile standard ale lui $S$ și $K$ în ipoteza de normalitate și cele două statistici $z$.
- 2. Calculați JB și ponderea care provine din boltire.
- 3. Decideți la nivelul de 5%.
- Raportați: două erori standard, două valori $z$, JB, decizia.

## Rezolvare

- 1. $\sqrt{6/4000} = 0{,}0387$, $\sqrt{24/4000} = 0{,}0775$; $z_S = -15{,}5$, $z_K = 154{,}9$
- 2. JB $= \frac{4000}{6}(0{,}36 + 36) = 240 + 24\,000 = 24\,240$, aproape în întregime din $K$
- 3. JB $> 5{,}99$: respingem normalitatea
- JB este suma celor două statistici $z$ la pătrat: $z_S^2 + z_K^2$.

înapoi Lista exercițiilor

# A8: momentele distribuției Student-$t$ [Propus]

Rezolvarea se discută la seminar

## Cerință

- $T \sim t(\nu)$; un model pentru randamentele zilnice este $r = m + s\,T$. Model: A7.
- 1. Calculați varianța și excesul de boltire pentru $t(5)$ și $t(10)$.
- 2. Găsiți $\nu$ pentru care excesul de boltire este 3.
- 3. Găsiți scala $s$ care dă o abatere standard zilnică de 1,2% pentru $\nu = 5$ și pentru $\nu = 6$.
- 4. Comparați $P(|X| > 4)$ pentru o $t(5)$ cu varianța 1 (valoarea din notebook) cu valoarea Normală.
- Raportați: patru momente, o valoare $\nu$, două valori ale scalei și un raport.

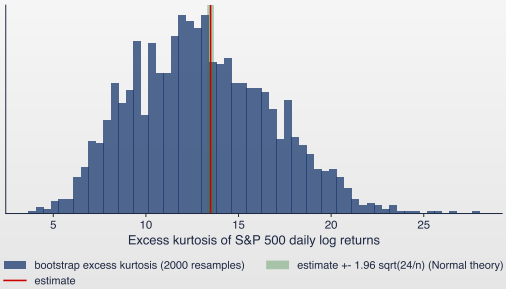
Înapoi Lista exercițiilor

# B1: cît de departe de normalitate este S&P 500? [Rezolvat]

- **Întrebarea**: cu ce precizie putem măsura cît de departe sînt randamentele logaritmice zilnice ale S&P 500 de distribuția Normală?
- **Date**: închiderile S&P 500, 2010–2026, randamente logaritmice zilnice în %
- **Cerințe**
  1. Calculați media, abaterea standard, asimetria, excesul de boltire și cea mai slabă zi, cu data ei.
  2. Calculați erorile standard ale lui $S$ și $K$ în ipoteza de normalitate și statistica JB.
  3. Aplicați bootstrap pentru $S$ și $K$, cu 2000 de reeșantionări ale zilelor, și rețineți percentilele 2,5% și 97,5%.
  4. Interpretare: de ce este intervalul bootstrap al lui $K$ mult mai larg decît cel din teoria Normală?
- **Raportați**: un tabel de momente, JB, două intervale pentru $K$, o frază
- Notebook: secțiunea B1

◂ Înapoi ▸ Lista exercițiilor

# B1: rezolvare [Rezolvat]




- $n = 4\,201$; media 0,045%, abaterea std. 1,08%; cea mai slabă zi $-12{,}8\%$, pe 16 martie 2020
- $S = -0{,}61$ (SE 0,038), $K = 13{,}49$ (SE 0,076); JB $= 32\,113$: normalitatea este respinsă
- Teoria Normală: $K \in [13{,}34; 13{,}64]$; bootstrap: $[6{,}7; 20{,}6]$; pentru $S$, bootstrap $[-1{,}54; 0{,}29]$
- Interpretare: $\sqrt{24/n}$ presupune date Normale; cu cozi groase, $K$ depinde de cîteva zile, iar prezența sau absența lor într-o reeșantionare îl modifică mult; abaterea standard bootstrap a lui $K$ este de 50 de ori mai mare
- Intervalul asimetriei conține 0: semnul lui $S$ nu este sigur

SFM_ch2_seminar

## B2: încă trei serii [Propus]

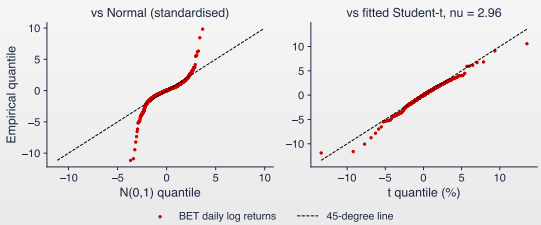
- **Întrebarea**: care dintre BET, Bitcoin și OMV Petrom este cel mai departe de distribuția Normală, cu și fără ziua sa cea mai extremă?
- **Date**: randamente logaritmice zilnice, 2010–2026 (Bitcoin din 2014), fiecare pe calendarul ei; model: B1
- **Cerințe**
  1. Calculați $n$, media, abaterea standard, $S$, $K$ și JB pentru fiecare serie.
  2. Eliminați cel mai mare randament în valoare absolută din fiecare serie și recalculați $K$.
  3. Interpretare: ce vă spune schimbarea lui $K$ despre boltire ca rezumat al riscului?
- **Raportați**: un tabel și două fraze
- Notebook: secțiunea B2

Înapoi Lista exercițiilor

# B3: distribuția Normală sau Student-$t$ pentru BET? [Rezolvat]

- **Întrebarea**: descrie o distribuție Student-$t$ randamentele zilnice ale BET mai bine decît distribuția Normală?
- **Date**: închiderile BET, 2010–2026, randamente logaritmice zilnice în %
- **Cerințe**
  1. Desenați QQ plot-ul randamentelor standardizate față de $N(0,1)$.
  2. Estimați o distribuție Student-$t$ prin verosimilitate maximă și desenați QQ plot-ul față de distribuția $t$ estimată.
  3. Comparați log-verosimilitățile și AIC ale celor două modele.
  4. Interpretare: unde greșește distribuția Normală?
  5. Interpretare: unde greșește distribuția $t$ estimată?
- **Raportați**: $\hat{\nu}$, $\hat{m}$, $\hat{s}$, două valori AIC, două QQ plots, o frază
- Notebook: secțiunea B3

# B3: rezolvare [Rezolvat]




- $\hat{\nu} = 2{,}96$, $\hat{m} = 0{,}078\%$, $\hat{s} = 0{,}628\%$; abaterea standard implicită $\hat{s}\sqrt{\hat{\nu}/(\hat{\nu}-2)} = 1{,}10\%$
- $\ell_t = -5496$ față de $\ell_N = -6224$; $\mathrm{AIC}_N - \mathrm{AIC}_t = 1454$: distribuția $t$ este clar preferată
- Cel mai mic randament $-11{,}9\%$; cuantila corespunzătoare: distribuția Normală $-3{,}9\%$, distribuția $t$ estimată $-13{,}4\%$
- Interpretare: distribuția Normală nu descrie niciuna dintre cozi (forma de S); distribuția $t$ descrie bine centrul și cea mai mare parte a cozilor, dar cuantilele ei extreme depășesc puțin datele

SFM_ch2_seminar

# B4: cîte zile de 4 sigma? [Propus]

- **Întrebarea**: care model estimează corect numărul zilelor de 4 sigma pentru S&P 500, DAX, BET și Bitcoin?
- **Date**: randamente logaritmice zilnice, 2010–2026 (Bitcoin din 2014); model: B3
- **Cerințe**
  1. Numărați zilele cu $|r_t - \bar{r}| > 4s$ din fiecare serie.
  2. Calculați numărul așteptat conform distribuției Normale și probabilitatea Poisson de a observa cel puțin numărul observat.
  3. Calculați numărul așteptat cu distribuția Student-$t$ estimată (același prag $\bar{r} \pm 4s$).
  4. Interpretare: are distribuția Student-$t$ cozi prea groase, prea subțiri sau potrivite pentru fiecare serie?
- **Raportați**: un tabel și o frază pentru fiecare serie
- Notebook: secțiunea B4

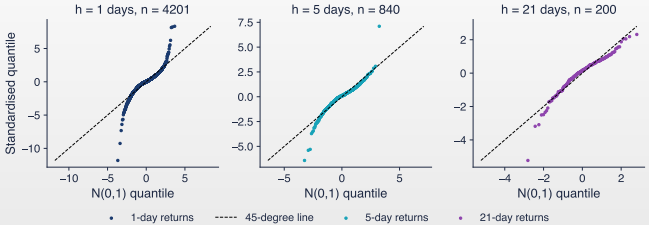
Înapoi Lista exercițiilor

# B5: devin randamentele S&P 500 Normale pe orizonturi mai lungi? [Rezolvat]

- **Întrebarea**: sînt randamentele pe 5 și pe 21 de zile ale S&P 500 mai apropiate de distribuția Normală decît cele zilnice?
- **Date**: randamentele logaritmice zilnice ale S&P 500, 2010–2026; randamentele pe $h$ zile sînt sume pe blocuri disjuncte de $h$ zile
- **Cerințe**
  1. Construiți randamentele logaritmice pe 5 și pe 21 de zile.
  2. Calculați $n$, $S$, $K$, eroarea standard $\sqrt{24/n}$ și JB cu p-value-ul lui pentru $h = 1, 5, 21$.
  3. Desenați cele trei QQ plots față de $N(0, 1)$.
  4. Interpretare: este Normală distribuția randamentelor lunare?
  5. Interpretare: de ce are testul mai puțină putere la $h = 21$?
- **Raportați**: un tabel cu trei rînduri, trei QQ plots, două fraze
- Notebook: secțiunea B5

◂ Înapoi ▸ Lista exercițiilor

# B5: rezolvare [Rezolvat]




| $h$ | $n$ | $S$ | $K$ | $\sqrt{24/n}$ | JB (p-value) |
|---|---|---|---|---|---|
| 1 | 4201 | −0,61 | 13,49 | 0,08 | 32 113 ($\approx 0$) |
| 5 | 840 | −0,62 | 7,06 | 0,17 | 1 797 ($\approx 0$) |
| 21 | 200 | −1,18 | 3,68 | 0,35 | 159 ($2{,}9 \times 10^{-35}$) |

- Interpretare: $K$ scade de la 13,49 la 3,68, dar randamentele lunare tot nu sînt Normale (asimetria −1,18, martie 2020); cu doar 200 de luni, eroarea standard a lui $K$ este 0,35, deci abaterile mai mici n-ar fi detectate

SFM_ch2_seminar

# B6: sînt independente randamentele BET? [Propus]

- **Întrebarea**: sînt randamentele zilnice ale BET independente, nu doar necorelate?
- **Date**: randamentele logaritmice zilnice ale BET, 2010–2026; model: B5
- **Cerințe**
  1. Calculați ACF pentru $r_t$ și pentru $|r_t|$ la lagurile 1–10 și banda $\pm 1{,}96/\sqrt{n}$.
  2. Calculați statistica Ljung–Box $Q(10)$ pentru $r_t$ și pentru $|r_t|$ și comparați cu $\chi^2_{0{,}95}(10)$.
  3. Interpretare: pot fi randamentele necorelate, dar nu independente?
- **Raportați**: două coloane ACF, două valori $Q$, o frază
- Notebook: secțiunea B6

înapoi Lista exercițiilor

# B7: seamănă Bitcoin cu un indice bursier? [Propus]

- **Întrebarea**: arată atît S&P 500, cît și Bitcoin efectul de levier și asimetria cîștig/pierdere?
- **Date**: randamente logaritmice zilnice, S&P 500 2010–2026, Bitcoin 2014–2026; model: B5
- **Cerințe**
  1. Calculați $L(k)$ (Noțiuni necesare azi (14/15)) pentru $k = 1, \ldots, 5$ și media lor, cu banda $\pm 1{,}96/\sqrt{n}$.
  2. Numărați zilele sub $\bar{r} - 4s$ și peste $\bar{r} + 4s$; testați dacă scăderile și creșterile sînt la fel de probabile (testul binomial, $p = 0{,}5$).
  3. Comparați $|q_{0{,}01}|$ cu $q_{0{,}99}$ și analizați asimetria.
  4. Interpretare: ce fapte stilizate ale indicilor bursieri are și Bitcoin?
- **Raportați**: un tabel și două fraze
- Notebook: secțiunea B7

◂ Înapoi ▸ Lista exercițiilor

# C1: devin mai subțiri cozile Bitcoin? [Propus]

- **Întrebarea**: a devenit mai subțire coada randamentelor zilnice ale Bitcoin din 2015, pe măsură ce piața s-a maturizat?
- **Date**: randamentele logaritmice zilnice ale Bitcoin și S&P 500, 2015–2026; modele: B3, B5
- **Cerințe**
  1. Estimați o distribuție Student-$t$ și calculați excesul de boltire pentru fiecare an calendaristic și fiecare serie.
  2. Calculați corelația rangurilor Spearman între an și $\hat{\nu}$.
  3. Estimați o distribuție $t$ pe 2015–2020 și una pe 2021–2026 și comparați $\hat{\nu}$.
  4. Propuneți o cale de a face răspunsul mai robust (intervale, alte ferestre, alte active cripto).
  5. Interpretare: ce ne pot spune 12 estimări anuale despre un trend?
- **Raportați**: un tabel, corelația rangurilor cu p-value-ul ei, un scurt plan de proiect
- Notebook: secțiunea C1

◂ Înapoi ▸ Lista exercițiilor

# C2: verificarea unui răspuns generat de AI [Propus]

- **Un student a întrebat un asistent AI: „Descrie distribuția randamentelor zilnice ale S&P 500 în 2010–2026." Răspunsul:**
- `(a) Excesul de boltire este 13,5; cum distribuția Normală are boltirea 0, boltirea S&P 500 este 13,5.`
- `(b) JB = 32 113 cu p = 0,000, ceea ce dovedește că randamentele urmează o distribuție Student-t.`
- `(c) În distribuția Normală, o zi de 4 sigma apare o dată la 15 787 de zile; cele 28 de zile de acest fel din 4 201 sînt doar ghinion.`
- `(d) Autocorelația la lagul 1 este doar −0,11, deci randamentele sînt independente și volatilitatea nu poate fi prevăzută.`
- `(e) Cu randamente logaritmice anuale N(7%, 18%^2), valoarea așteptată după un an a unei investiții de 100 este 100 exp(0,07) = 107,25.`
- **Cerințe**
  1. **Pentru fiecare afirmație precizați dacă este corectă; dacă nu, formulați afirmația corectă și dați valoarea corectă din notebook (secțiunea C2).**
- **Raportați: o listă de cinci verdicte, fiecare cu un rînd de justificare.**

Înapoi Lista exercițiilor

## Idei de reținut

- Distribuția Normală face zilele de 4 sigma aproape imposibile; piețele reale au o asemenea zi la fiecare cîteva luni
- Prețuri lognormale: mediana este sub medie, iar diferența provine din volatility drag
- Asimetria și boltirea sînt fragile: raportați intervale bootstrap și extremele
- Distribuția Student-$t$ descrie datele mult mai bine decît distribuția Normală, dar nu perfect
- Randamentele sînt aproape necorelate, dar nu independente: mărimea lor este previzibilă
- Un răspuns generat de AI este o ciornă: verificați fiecare definiție, fiecare model și fiecare valoare numerică

## După seminar

- Cursul 2 dezvoltă fiecare temă de azi: modelele Normal și lognormal, CLT, momente și teste, distribuția Student-$t$ și șase fapte stilizate
- C1 poate deveni un proiect de echipă: ferestre mobile, intervale bootstrap, mai multe active cripto
- Lectură: Franke, Härdle & Hafner (2019), secț. 3.3 și secț. 11.2; exerciții rezolvate: Borak, Härdle & López-Cabrera (2013)

# Bibliografie


- Borak, S., Härdle, W.K., López-Cabrera, B. (2013). *Statistics of Financial Markets: Exercises and Solutions*, 2nd ed. Springer.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). *Statistics of Financial Markets: An Introduction*, 5th ed. Springer.
- Jarque, C.M., Bera, A.K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172.
- Ljung, G.M., Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.
- Student (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25.