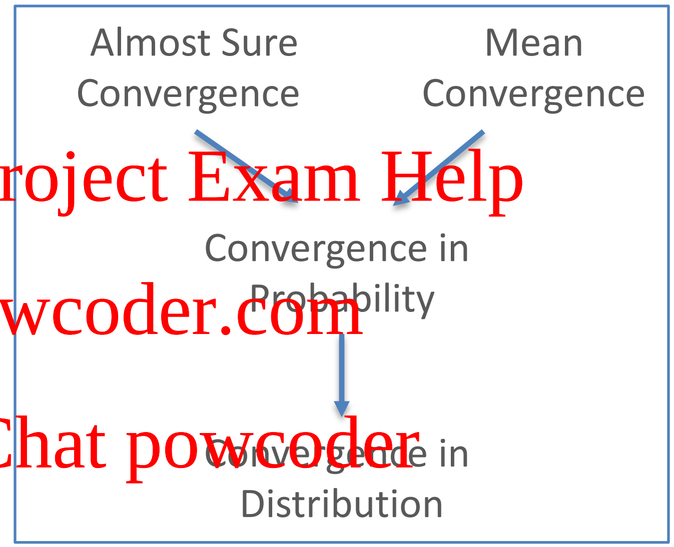
Economics 403A

Sampling Distributions and Limits

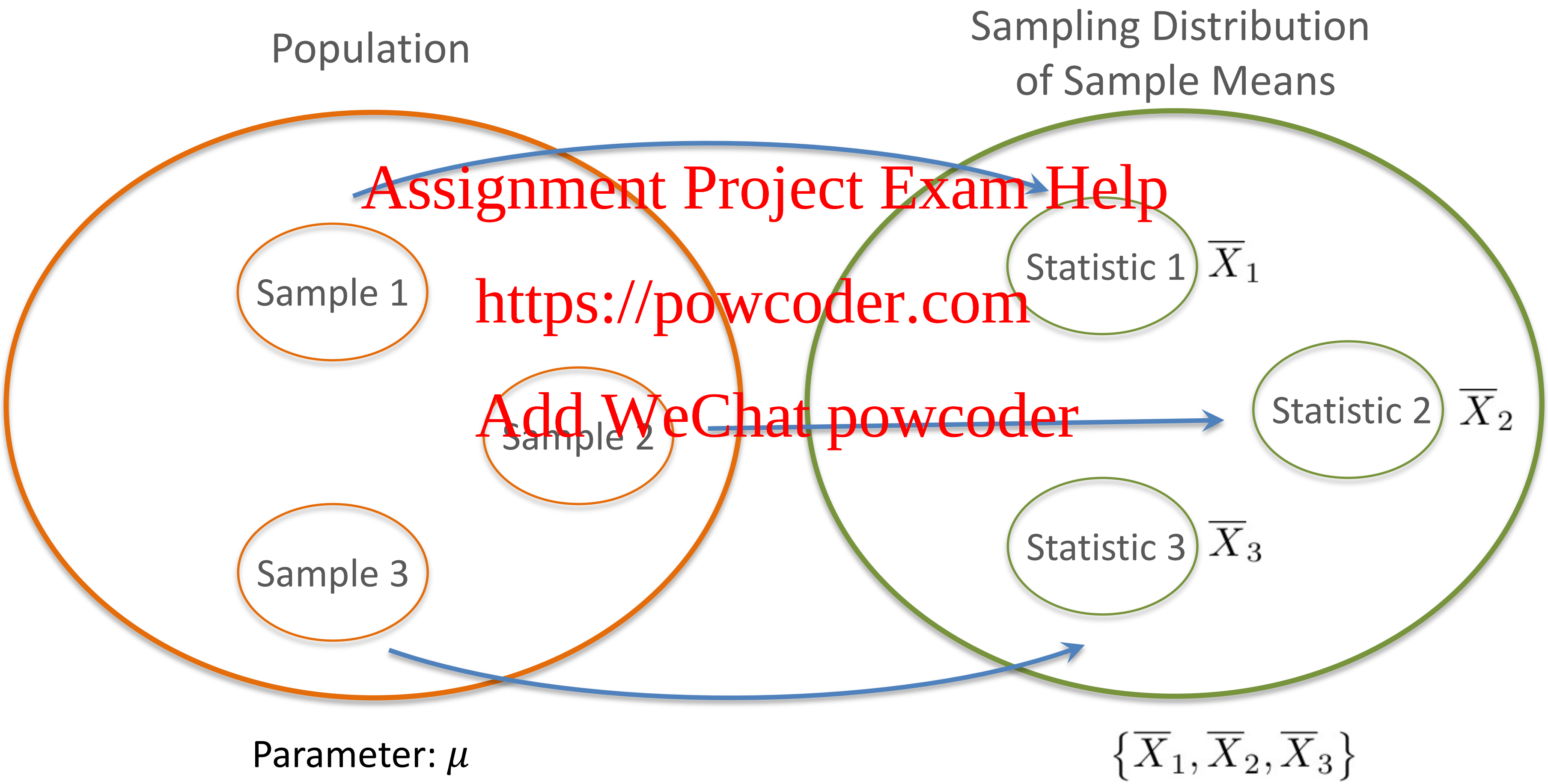
Dr. Randall R. Rojas

Today's Class

- Sampling Distributions
- Limits
  - Convergence in Probability
  - Weak Law of Large Numbers
  - Convergence in Probability 1
  - Strong Law of Large Numbers
  - Convergence in Distribution
  - Central Limit Theorem
- Monte Carlo Approximations
- Normal Distribution Theory
  - Chi-Squared Distribution
  - t-Distribution
  - F-Distribution



Sampling Distributions



Sampling Distributions

Anatomy of a Basic Sampling Algorithm

for i= 1:k

1. Choose a sample size n, and a statistic

2. Randomly draw a sample from the population with the same size

3. Estimate the statistic from the sample

end

Sampling Distributions

Def: Sampling Variability

= The variability among random samples from the same population.

Def: Sampling Distribution

= A probability distribution that characterizes some aspect of sampling variability.

= The distribution of a statistic given repeated sampling.

A sampling distribution tells us how close the resemblance between the sample and population is likely to be

Goal: We use probability theory to derive a distribution for a statistic, which allows us to make statistical inferences about population parameters.

Sampling Distributions

Suppose $X_1, X_2, \ldots, X_n$ is a sample from some distribution, f is a function of this sequence, and we want to estimate the distribution of the r.v., $Y = f(X_1, X_2, \ldots, X_n)$, i.e., Y's 'sampling distribution'.

Q: How can we obtain the distribution of Y?

A: Analytically (works well when n is small) or Approximately (through sampling)

Sampling Distributions

- Example: Given a sample X_1, X_2 of size $n = 2$, with p.m.f. p_X, and $Y_2 = (X_1 X_2)^{1/2}$, find the distribution of Y_2.

$$p_X(x) = \begin{cases} 1/2 & x = 1 \\ 1/4 & x = 2 \\ 1/4 & x = 3 \\ 0 & \text{otherwise.} \end{cases}$$

Note: Y_2 is known as the ***Geometric Average***, which is widely used in finance.

- Solution: We can compute it analytically

y	Sample	$p_{Y_2}(y)$
1	$\{(1,1)\}$	$(1/2)(1/2) = 1/4$
$\sqrt{2}$	$\{(1,2),(2,1)\}$	$(1/2)(1/4) + (1/4)(1/2) = 1/4$
$\sqrt{3}$	$\{(1,3),(1,3)\}$	$(1/2)(1/4) + (1/4)(1/2) = 1/4$
2	$\{(2,2)\}$	$(1/4)(1/4) = 1/16$
$\sqrt{6}$	$\{(2,3),(3,2)\}$	$(1/4)(1/4) + (1/4)(1/4) = 1/8$
3	$\{(3,3)\}$	$(1/4)(1/4) = 1/16$

Sampling Distributions

- Application: Multiple-Period Realized Return

Suppose you invest \$1000 at the beginning of year 1, and get a 12% return at the end this year, and then get 8% at the end of year 2. What was your 'average' return over the two years?

Option 1: Arithmetic Average: (12+8)/2 = 10%

Q: Is this correct? No

Q: Why is this the wrong answer? A: Does not take into account interest over interest.

- If you start with \$1000, then at the end of 2 years you have \$1000 (1.12)(1.08) = **\$1209.6**.
- The amount of money you would get if you invested \$1000 at 10%/year for two years is $1000*(1.1)^2$ = \$1210. They are not the same!

Option 2: Geometric Average: $[(1+0.12)(1+0.08)]^{1/2} - 1 = 0.09982$=9.982%

The geometric average gives us the correct per annum return: $1000(1.09982)^2$=**\$1209.6**.

Multiple-Period Realized Return
Arithmetic vs. Geometric Average

- Example: Time Series of HPR for the S&P 500

Period	HPR (decimal)	Gross HPR =1+HPR
2001	-0.1189	0.8811
2002	-0.2210	0.7790
2003	0.2869	1.2869
2004	0.1088	1.1088
2005	0.0491	1.0491

Arithmetic Average:

$= 1/5(-0.1189 + \ldots + 0.0491)$

$= 0.0210$

$= 2.10\%$

Geometric Average:

$= (0.8811 \times \ldots \times 1.0491)^{1/5} - 1$

$= 0.0054$

$= 0.54\%$

The larger the swings in rates of return, the greater the discrepancy between the arithmetic and geometric averages.

Multiple-Period Realized Return
Arithmetic vs. Geometric Average

- Example: S&P 500 annual returns from 2000-2010
- Q: Given the two averages, if you had invested $10,000 in the S&P500 from 2000-2010, how much would you have in your account?
  - R={-9.2, -11.9, -22.1, 28.7, 10.9, 4.9, 15.8, 5.5, -37.0, 26.5, 15.1}

- Arithmetic Avg = 2.47%,   Geometric Avg = 0.41%

  $10,247.00 $10,041.00 } correct amount

- Reality?

  After adjusting for inflation
  → Arithmetic ($9,994), Geometric ($9,798)

Multiple-Period Realized Return Arithmetic vs. Geometric Average

Concluding Remarks

- Volatility lowers investment returns.
- Most returns are reported as an arithmetic average because this is the highest average that can be reported.
  - Arithmetic averages are accurate only if there is no volatility.
- Careful with e.g., mutual fund and hedge fund managers! –They usually quote arithmetic averages.

Sampling Distributions

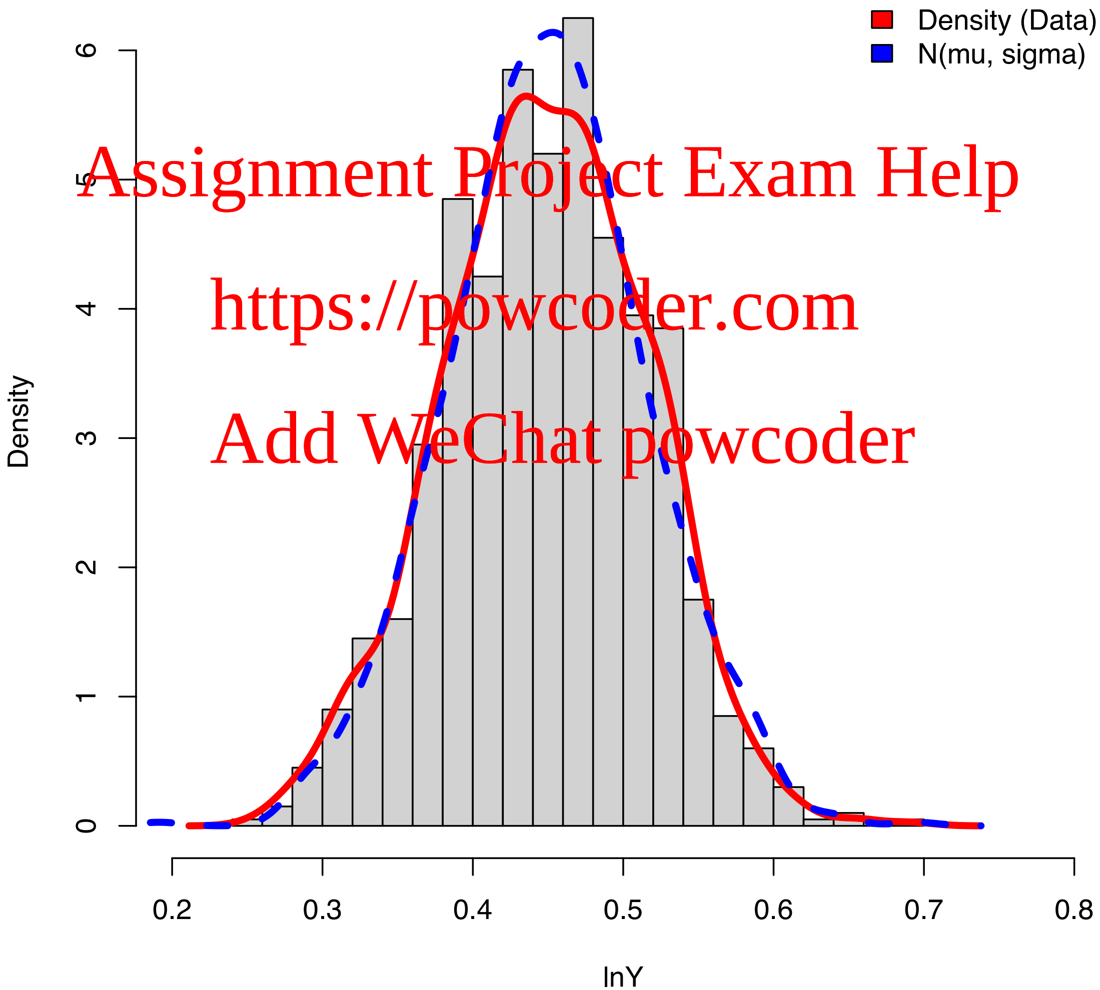
- Example: Given a sample $X_1, X_2, \dots, X_{20}$ of size $n = \mathbf{20}$, with p.m.f. p_X, and $Y_2 = (X_1 \cdots X_{20})^{1/2}$, find the distribution of Y_2.

- Note: We ***cannot*** compute it analytically! Even computationally it would be difficult because of the large number of operations.

Q: How do we find Y_2?

A: Approximation: $\ln Y = \frac{1}{n} \sum_{i=1} nX_i$

Sampling Distributions



Limits

Convergence in Probability

A sequence of random variables $X_1, X_2, X_3, \ldots X_n$ converges in probability to a random variable X, i.e., $X_n \xrightarrow{p} X$ if:

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0, \quad \forall \varepsilon > 0$$

Example: Let $X_n \sim Exp(n)$, show that $X_n \xrightarrow{p} 0$. That is, the sequence $X_1, X_2, X_3, \ldots X_n$ converges in probability to the zero random variable X.

→Proof:

$$\lim_{n\to\infty} P\left(|X_n - 0| \geq \varepsilon\right) = \lim_{n\to\infty} P\left(X_n > \varepsilon\right)$$

$$= \lim_{n\to\infty} e^{-n\varepsilon}$$

$$= 0, \ \forall \varepsilon > 0$$

Limits

Convergence in Probability

(Textbook Example)

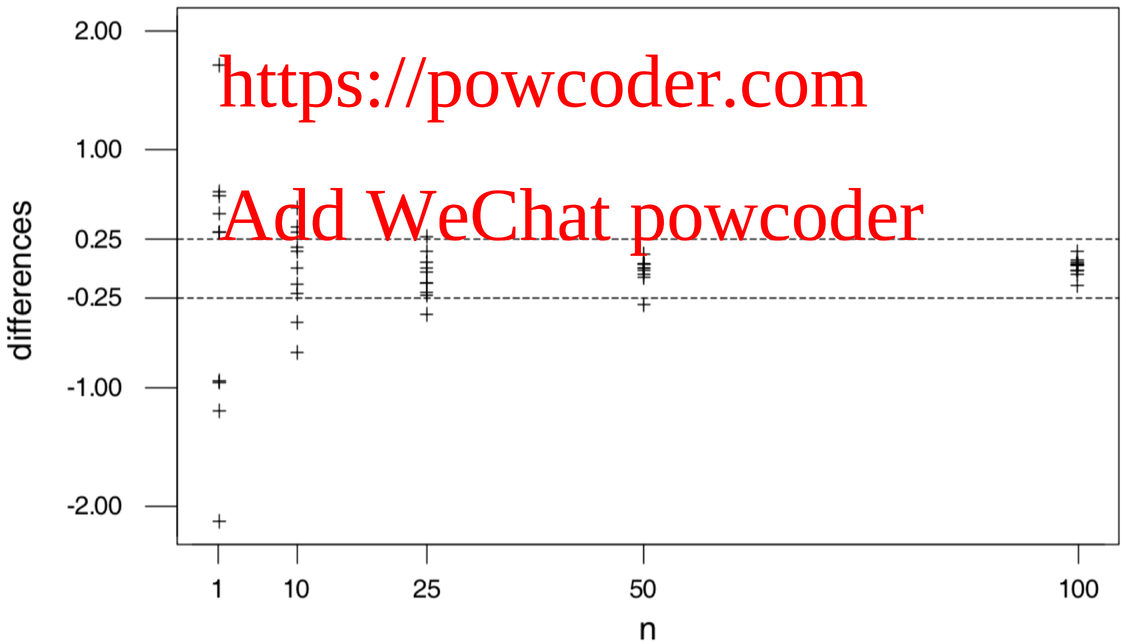


Figure 4.2.1: Plot of 10 replications of $\{X_n - Y\}$ illustrating the convergence in probability of X_n to Y.

Limits

Weak Law of Large Numbers (WLLN)

Let $X_1, X_2, X_3, \ldots X_n$ be i.i.d random variables with a finite expected value $E[X_i] = \mu < \infty$. Then for any $\varepsilon > 0$,

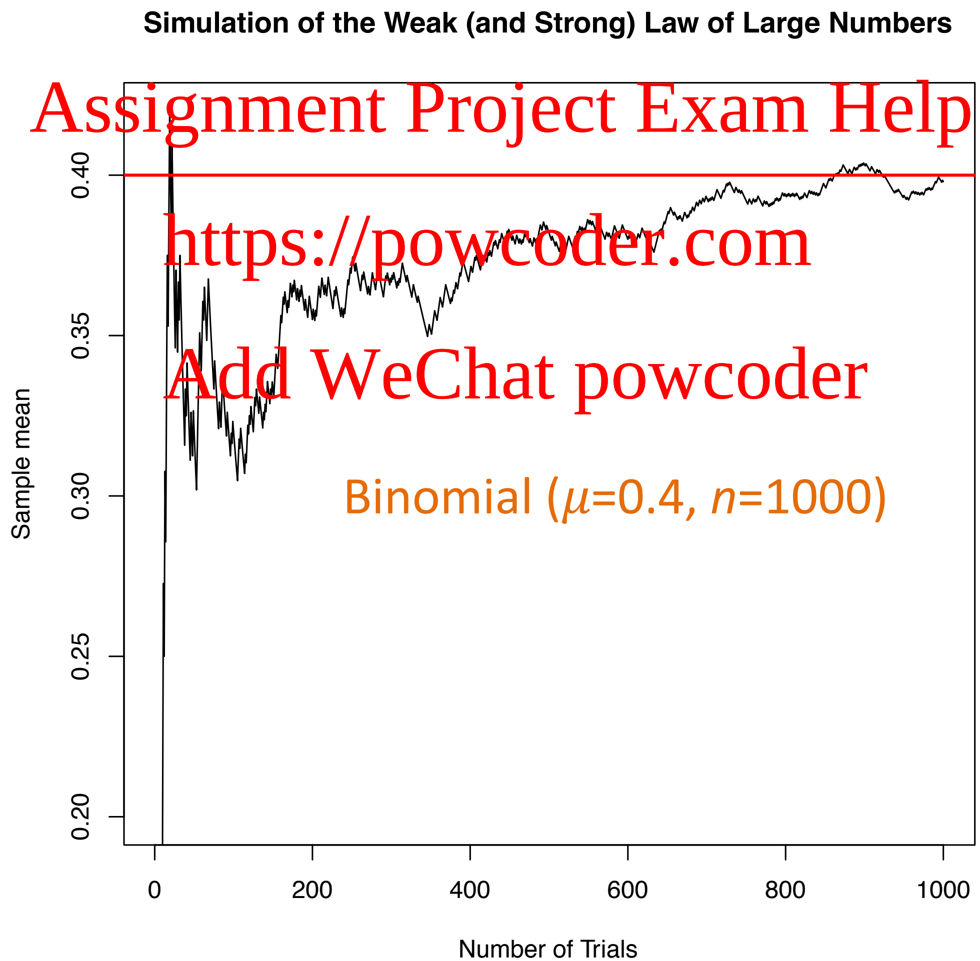
$$\lim_{n \to \infty} P\left(|\overline{X} - \mu| \geq \varepsilon\right) = 0.$$

Note: A common notation is $M_n = \overline{X}$.

The WLLN is also known as **Bernoulli's Theorem**. In simpler terms,it states that: "*the mean of the results obtained from a large number of trials is close to the population mean*".

Limits

Weak Law of Large Numbers (WLLN)

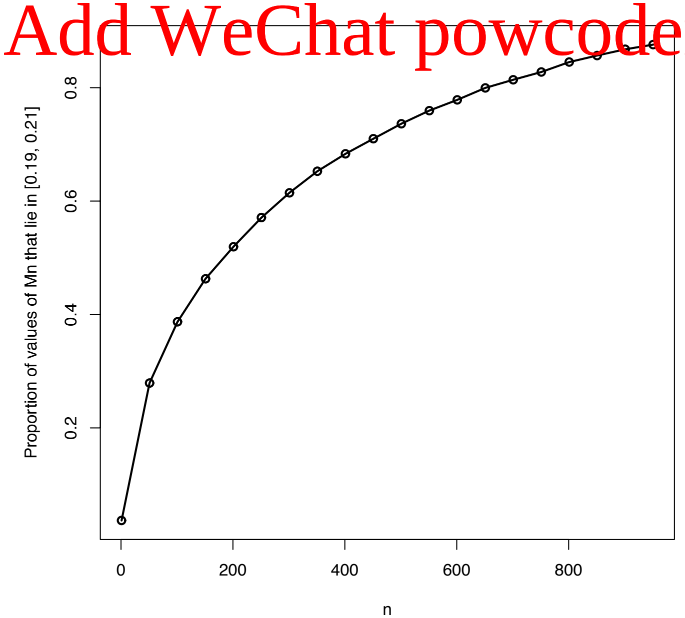


Limits

Weak Law of Large Numbers (WLLN)

Example (4.2.12): Generate i.i.d. $X_1, \ldots, X_n$ distributed Exponential(5) and compute M_n when $n = 20$. Repeat this N times, where N is large (if possible, take $N = 10^5$, otherwise as large as is feasible), and compute the proportion of values of M_n that lie between 0.19 and 0.21. Repeat this with $n = 50$.

→Since $X \sim Exp(5)$, $\mu = 1/5 = 0.2$. Therefore, as n increases, the proportion of values that lie between 0.19 and 0.21 should increase.



Limits

Convergence in Probability 1 (“almost surely”)

A sequence of random variables $X_1, X_2, \ldots X_n$ converges *almost surely* to a random variable X, i.e., $X_n \xrightarrow{a.s.} X$, if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Example: A fair coin is tossed once. The sample space is S={H,T}. We define a sequence of random variables $X_1, X_2, \ldots X_n$ on this sample space according to:

$$X_n(s) = \begin{cases} \frac{n}{n+1} & \text{if s=H} \\ (-1)^n & \text{if s=T} \end{cases} \rightarrow \begin{array}{l} s = H \to X_n = \{1/2, 2/3, \ldots\} \to 1 \text{ as } n \to \infty \\ s = T \to X_n = \{-1, 1, \ldots\} \end{array}$$

(does ***not*** converge)

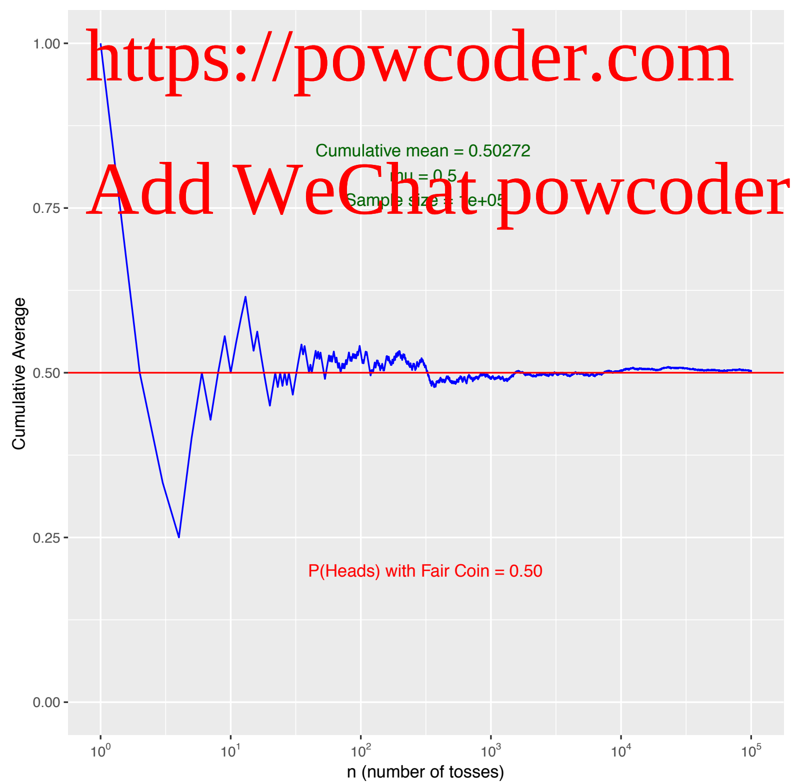
$$P\left(\lim_{n\to\infty} X_n = 1\right) = P(H) = 1/2$$

X_n converges if S=H but not if s=T, but since the probability that X_n converges to X is equal to 1 as $n \to \infty$, then $X_n \xrightarrow{a.s.} X$

Limits

Strong Law of Large Numbers

Let $X_1, X_2, \dots X_n$ be i.i.d., random variables each having a finite mean $E[X_i] = \mu < \infty$, and $M_n = \frac{X_1 + \cdots + X_n}{n}$, then $M_n \xrightarrow{a.s.} \mu$.



Limits

Strong Law of Large Numbers

Application: ***Business Growth***

According to the LLN as additional units are added to a sample, the average of the sample converges to the average of the population.

→The LLN implies that the more a company grows, the harder it is for the company to sustain that percentage of growth.

Company A: $100M (Market Cap), year 1 growth rate = 50%.

- A 50% growth rate for the next 6 years would grow the company from $100M to $1.2B.

Company B: $20M (Market Cap) year 1 growth rate = 50%.

- A 50% growth rate for the next 6 years would grow the company from $20M to $228M

Which company has more room for growth (stock appreciation)?

Limits

Convergence in Distribution

A sequence of random variables $X_1, X_2, \dots X_n$ converges in *distribution* to a random variable X, i.e., $X_n \xrightarrow{d} X$, if $\lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x)$ for all x at which $F_X(x)$ is continuous.

Example: Let $X_1, X_2, \dots X_n$ be a sequence of r.v's such that

$$F_{X_n}(x) = \begin{cases} 1 - \left(1 - \frac{1}{n}\right)^{nx} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Show that X_n converges in distribution to Exp(1).

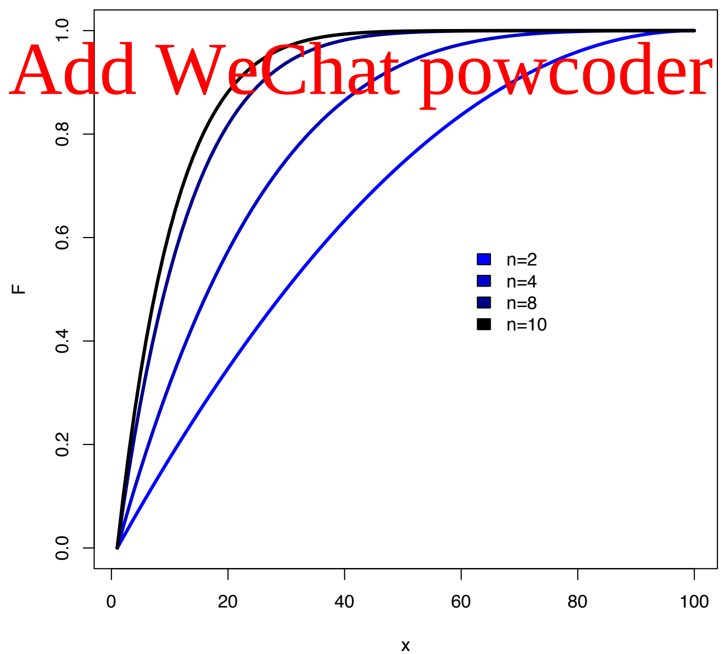
Limits

Convergence in Distribution

Solution: For $x \leq 0$, $F_{X_n}(x) = F_X(x) = 0$, for $n = 2,3, \ldots$

For $x \geq 0$, $\lim\limits_{n\to\infty} F_{X_n}(x) = \lim\limits_{n\to\infty} \left(1 - \left(1 - \frac{1}{n}\right)^{nx}\right) = \lim\limits_{n\to\infty} 1 - e^{-x} = F_X(x), \quad \forall x$

$\therefore X_n \xrightarrow{d} X$



Limits

Central Limit Theorem

Let $X_1, X_2, \ldots X_n$ be i.i.d r.v's with $E[X_i] = \mu < \infty$ and $0 < Var[X_i] = \sigma^2 < \infty$. Then, the r.v.

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the Standard Normal ($N(0,1)$) random variable, $\Phi(x)$ as $n \rightarrow \infty$, i.e.,

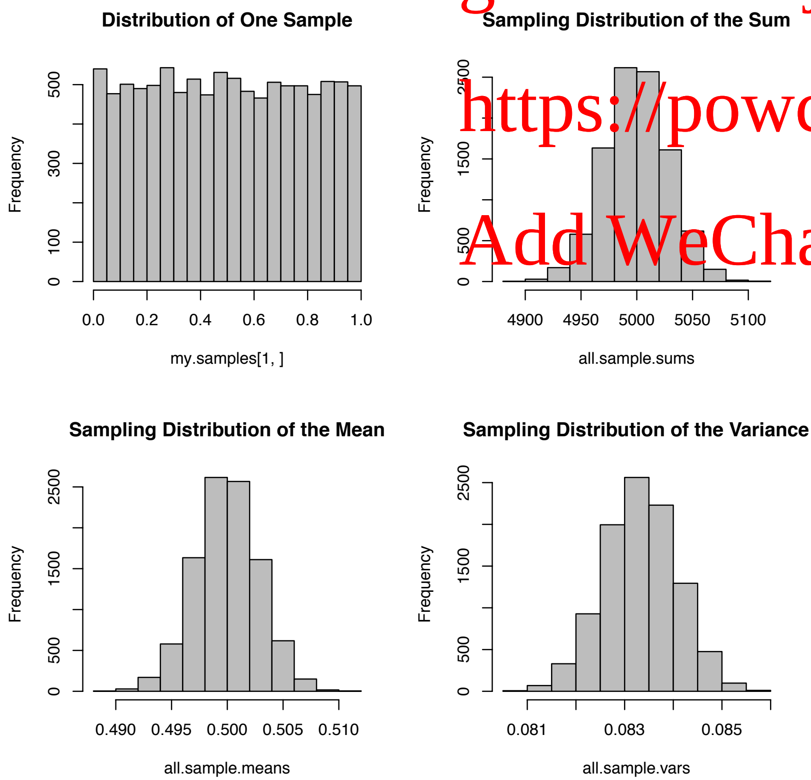
$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x), \quad \forall x \in \mathbb{R}$$

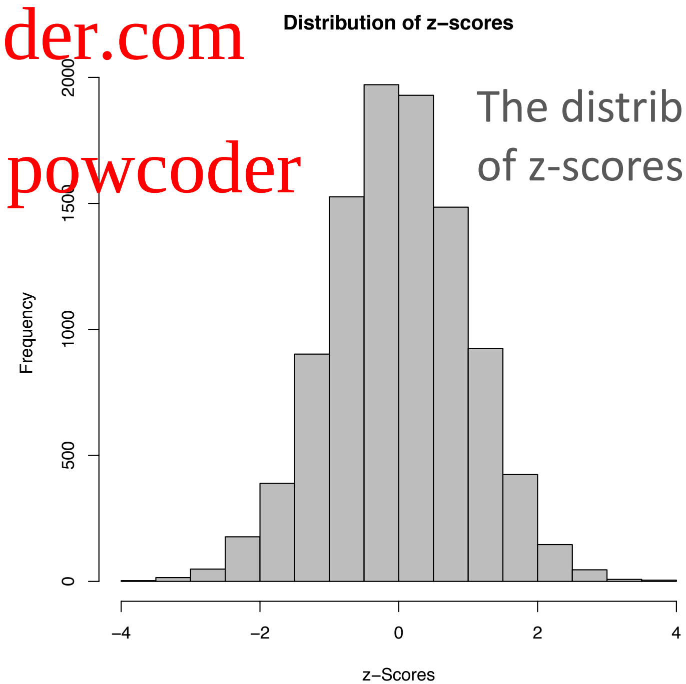
Note: $\Phi(x)$ is the standard normal cdf.

Limits

Central Limit Theorem

Example: n= 1000 samples from a Unif[0, 1]





The distribution of z-scores $\sim N(0,1)$

Limits

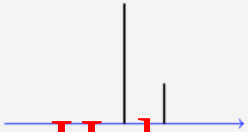
Assumptions:

- $X_1, X_2 \dots$ are iid Bernoulli(p).
- $Z_n = \dfrac{X_1 + X_2 + \dots + X_n - np}{\sqrt{np(1-p)}}$.

We choose $p = \frac{1}{3}$.

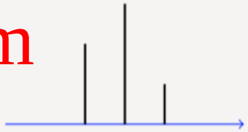
$$Z_1 = \frac{X_1 - p}{\sqrt{p(1-p)}}$$

PMF of Z_1



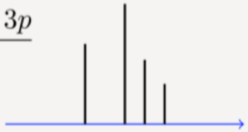
$$Z_2 = \frac{X_1 + X_2 - 2p}{\sqrt{2p(1-p)}}$$

PMF of Z_2



$$Z_3 = \frac{X_1 + X_2 + X_3 - 3p}{\sqrt{3p(1-p)}}$$

PMF of Z_3



$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - 30p}{\sqrt{30p(1-p)}}$$

PMF of Z_{30}

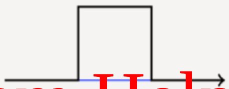
Limits

Assumptions:

- $X_1, X_2 \dots$ are iid Uniform(0,1).
- $Z_n = \dfrac{X_1 + X_2 + \dots + X_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$.

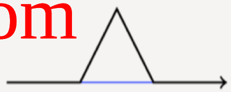
$$Z_1 = \frac{X_1 - \frac{1}{2}}{\sqrt{\frac{1}{12}}}$$

PDF of Z_1



$$Z_2 = \frac{X_1 + X_2 - 1}{\sqrt{\frac{2}{12}}}$$

PDF of Z_2



$$Z_3 = \frac{X_1 + X_2 + X_3 - \frac{3}{2}}{\sqrt{\frac{3}{12}}}$$

PDF of Z_3

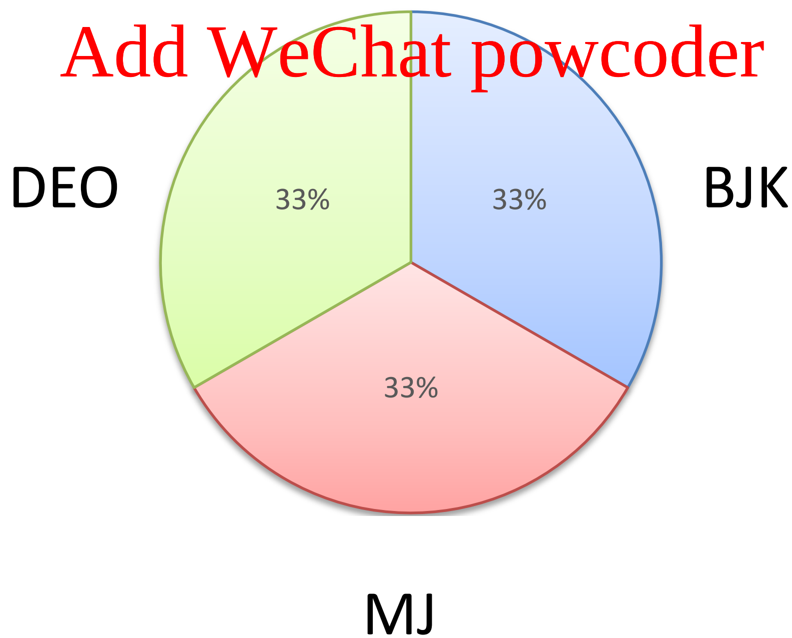
$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - \frac{30}{2}}{\sqrt{\frac{30}{12}}}$$

PDF of Z_{30}

Limits

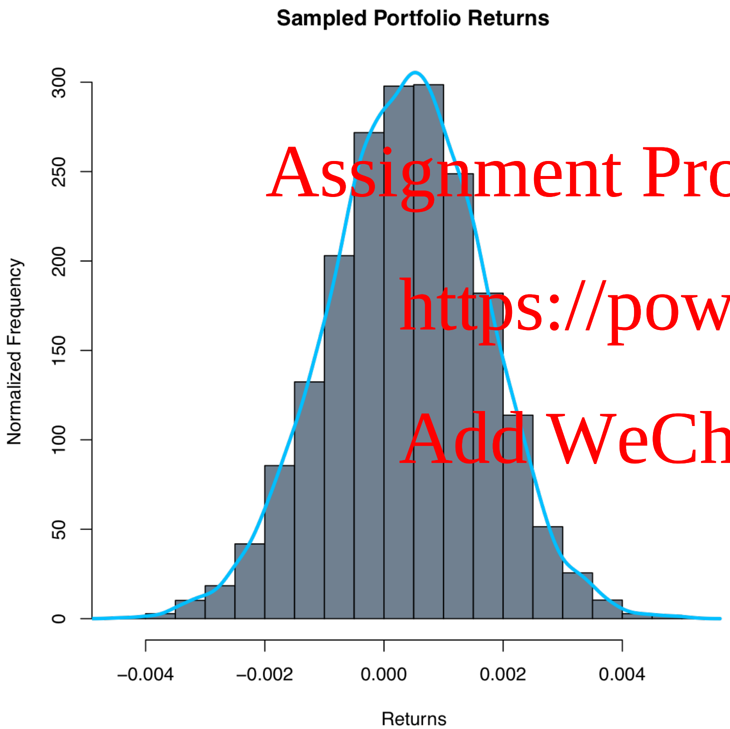
- Application: Investors of all types rely on the CLT to analyze stock returns, construct portfolios and manage risk.

- Example: Portfolio with 3 ETFs all equally weighted

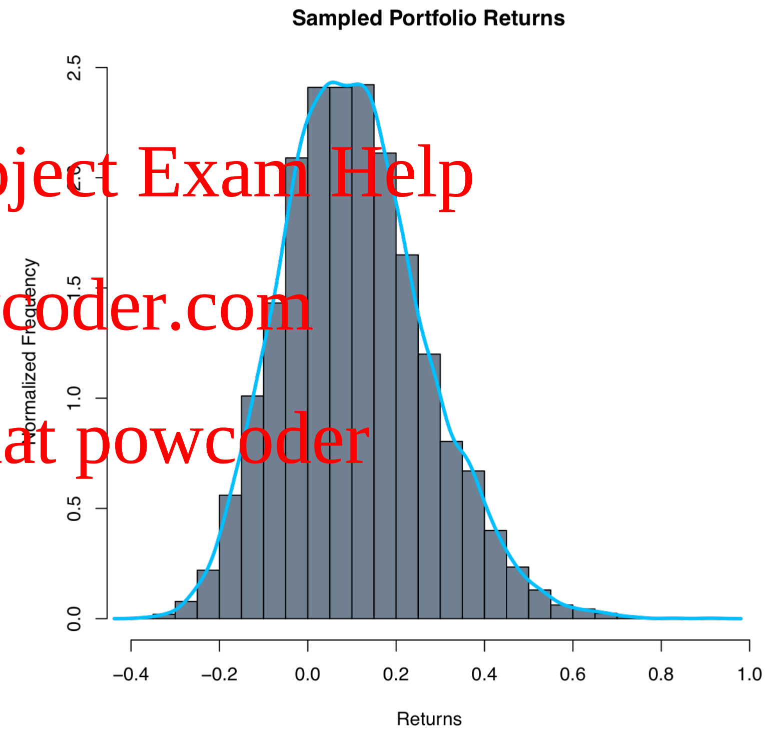


Limits

From Jan 1st, 2016 to September 1st, 2018



Daily



Annual

Note: See Portfolio.R

Monte Carlo Approximations

- Q: If X is a r.v where $X \sim f(X)$ then $g(X)$ is also a r.v., but what is the probability distribution of $g(X)$?

Example: Let $X \sim f(X) = \text{Unif}[1,4]$ and $Y = g(X) = (X-3)^2$, find $E[Y]$.

Step 1: $p(X) = 1/4 \ \forall\ X$ values in {1,2,3,4}

Step 2: $g(X)$

X = 1 → g(1) = 4

X = 2 → g(2) = 1

X = 3 → g(3) = 0

X = 4 → g(4) = 1

Step3: $P(g(X)) = P(Y)$

$$Y = 0, 1, 4$$

$$P(Y=0) = 1/4, P(Y=1) = 1/2, P(Y=4) = 1/4$$

Step 4: $E[g(X)] = \sum Y\, P(Y)$

$$= 0(1/4) + 1(1/2) + 4(1/4)$$

$$= 3/2$$

Monte Carlo Approximations

- Q: What if we don't't' know $P(Y)$? Do we need it?
- Example: Let $X \sim f(X) = \text{Unif}[1,4]$ and $Y = g(X) = (X-3)^2$, find $E[Y]$.

$$E[Y] = E[g(X)] = \sum Y\,P(Y) = \sum g(X)\,P_X(X)$$

$$= (1-3)^2\,P(X=1) + (2-3)^2\,P(X=2) + (3-3)^2\,P(X=3) + (4-3)^2\,P(X=4)$$

$$= 4\,(1/4) + 1(1/4) + 0(1/4) + 1(1/4)$$

$= 3/2 \rightarrow$ same as before without knowing $P(Y)$ ☺

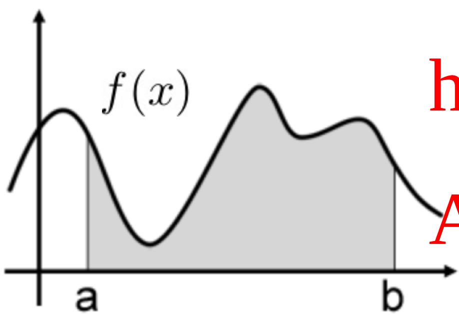
Conclusion: We do not need to know $P(Y)$ in order to compute $E[Y]$ provided we know $P(X)$.

= ***Foundation of MC Approximations!***

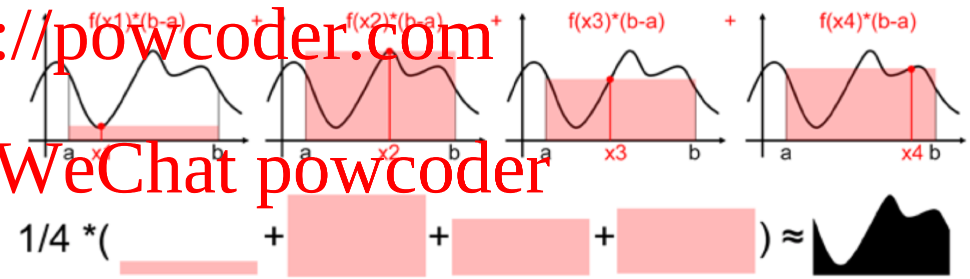
Monte Carlo Approximations

Application: Approximating Integrals

- How can we estimate: $F = \int_a^b f(x)\,dx$?



→



$$\langle F^N \rangle = (b-a)\frac{1}{N}\sum_{i=0}^{N-1} f(X_i)$$

Choose the points x at random between a and b, evaluate $f(x)$→ compute the area $A = l \times w$

$$\rightarrow \; P(\lim_{x\to\infty} \langle F^N \rangle = F) = 1 \quad \rightarrow \; E\left[\langle F^N \rangle\right] = F$$

(Converges in probability) (by the LLN)

Algorithmic Implementation of Sequential Bayesian Learning

Application: Approximating Integrals

Generalization (arbitrary PDF)

Since $F = \int_D f(\vec{x})d\vec{x} = \int_D \frac{f(\vec{x})}{p(\vec{x})}p(\vec{x})d\vec{x}$ for any PDF p

but $\int_D \frac{f(\vec{x})}{p(\vec{x})}p(\vec{x})d\vec{x} = E\left[\frac{f(\vec{x})}{p(\vec{x})}\right], x \sim p$

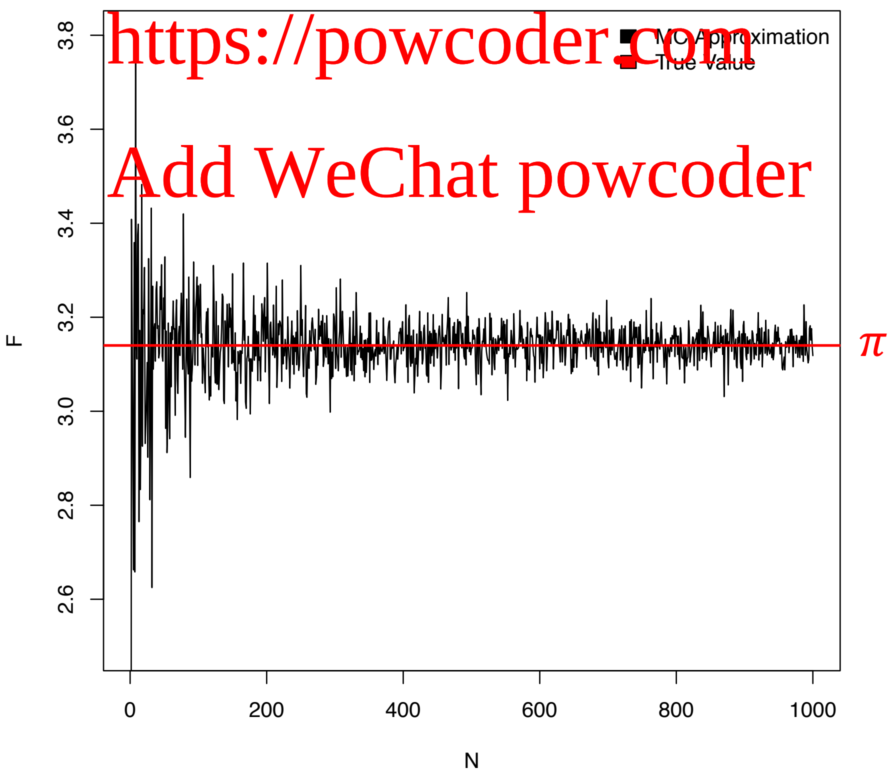
then $F_N = \frac{1}{N}\sum_{i=1}^{N}\frac{f(\vec{x}_i)}{p(\vec{x}_i)}$.

→ (by the LLN) as $\boxed{N \to \infty, F_N \xrightarrow{a.s} E\left[\frac{f(\vec{x})}{p(\vec{x})}\right] = F}$

Monte Carlo Approximations

- Example: Approximate the integral: $F = \int_0^1 4\sqrt{1-x^2}\,dx = \pi$

Let $X_1, X_2, \ldots, X_n$ = i.i.d r.v.s. ~ unif[0,1].$\rightarrow F^1 = f(X_1), \ldots, F^N = f(X_N)$ are i.i.d. r.v's with mean $\langle F^N \rangle$ and $E[\langle F^N \rangle] = F = \pi$



For $N = 1000$
$\rightarrow E[\langle F^N \rangle] = 3.15$

For $N = 10^5$
$\rightarrow E[\langle F^N \rangle] = 3.14$

Monte Carlo Approximations

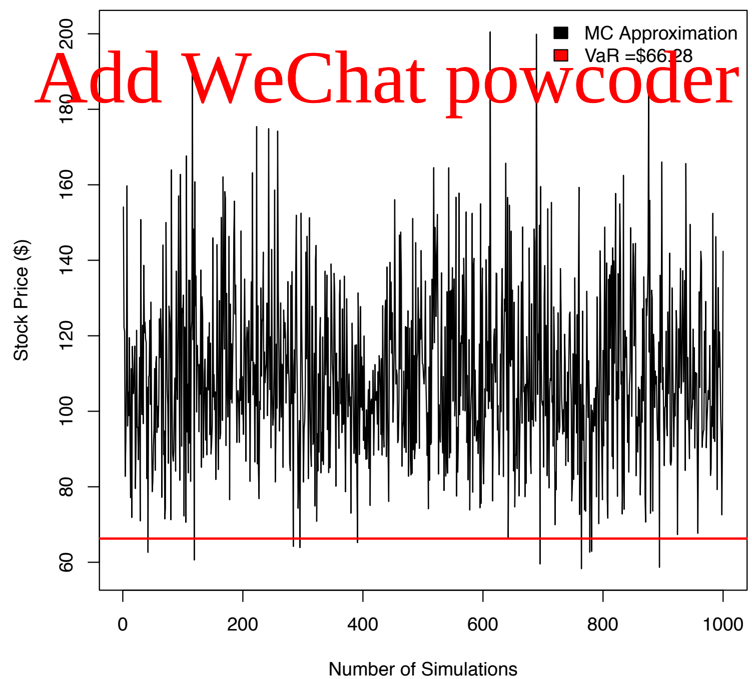
- Example: Value at Risk

  Given a current stock price (of your choice), what is the lowest it can go (worst-case scenario) 20 days from now with 99% probability of occurrence?

- Model: Geometric Brownian Motion → $S_{t+1} = S_t \times (1 + \mu \Delta t + \sigma \varepsilon \sqrt{t})$
  - Standard model for stock price movements

Assumptions:

S_0 = \$100
σ = 0.1 (10% drift)
μ = 0.2 (20% volatility)
ε ~unif[0,1]



VaR = \$66.28

Normal Distribution Theory

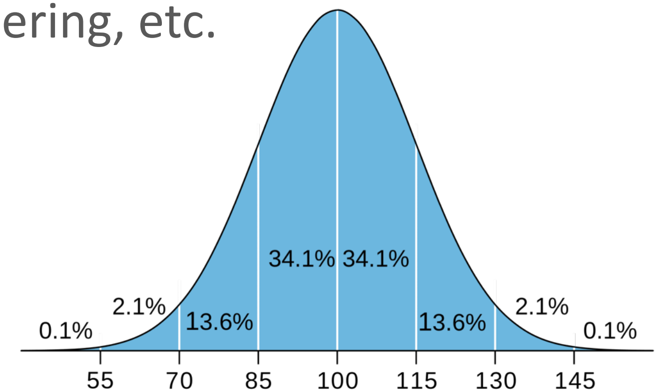
Normal Distribution $N(\mu,\sigma)$

- Suppose $X_i \sim N(\mu, \sigma_i^2)$, for $i = 1,2, \ldots, n$ and that they are independent r.v's. Let $Y = (\sum_i a_i X_i) + b$ for some constants $\{a_i\}$ and b. Then

$$Y \sim N\left(\sum_i a_i\mu_i + b, \sum_i a_i^2\sigma_i^2\right).$$

Applications: finance models (e.g., mean variance analysis), business, economics, science, engineering, etc.

PDF: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; -\infty < x < +\infty$



Normal Distribution Theory

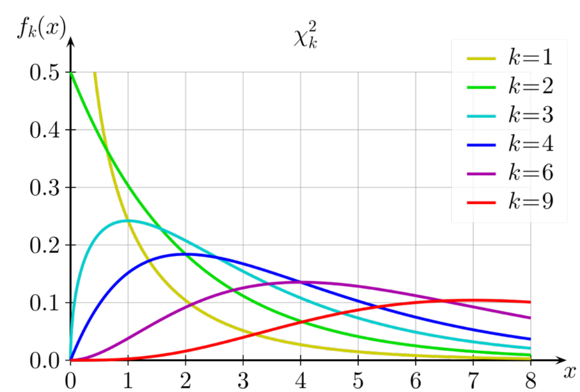
χ^2 –Distribution

- The *chi-squared* distribution with ν degrees of freedom (or $\chi^2(\nu)$) is the distribution of the random variable $Z = X_1^2 + X_2^2 + \cdots + X_\nu^2$, where $X_1, \ldots, X_n$ are i.i.d., each with the standard normal distribution $N(0,1)$.

- Applications: population variance estimation, goodness-of-fit tests, contingency tables, etc.

- PDF: $f(x) = \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2}, x > 0,$

  where, $\mu = \nu, \sigma^2 = 2\nu, M(t) = (1-2t)^{-\nu/2}$



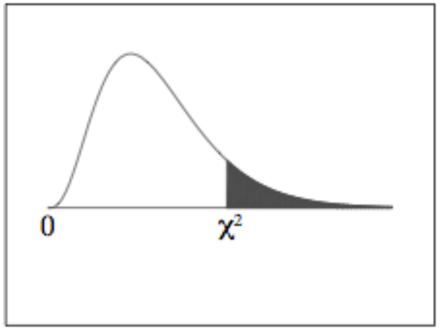
χ^2-Distribution

Examples

1. Find the value of χ^2_5 such that the area on the right tail is equal to 0.05.

Solution: df = 5, and α = 0.05

→ χ^2_5 = 11.070



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750

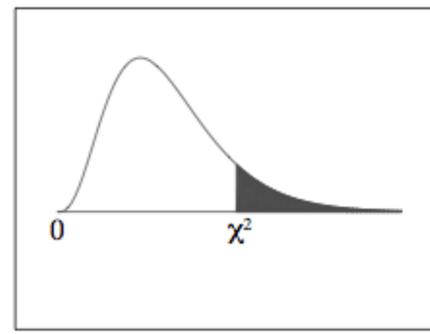
χ^2-Distribution

Examples

2. Find the value of χ^2_3 that represents the 97.5th percentile.

Solution: df = 3, and α = 0.025

→ $\chi^2_3 = 9.348$



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750

χ^2-Distribution

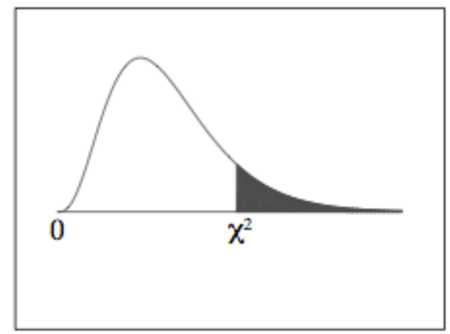
Examples

3. Given $X \sim \chi^2_4$, find P(0.484 < X < 7.779).

Solution: df = 4.

→P(0.484 < X < 7.779) = 0.975-0.100

= 0.874



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750

Normal Distribution Theory

t-Distribution

- The *t-distribution* with *n* degrees of freedom (or Student-t), is the distribution of the random variable

$$Z = \frac{X}{\sqrt{\left(X_1^2 + \cdots X_n^2\right)/n}}$$

where $X_1, X_2, \ldots, X_n$ are i.i.d., each with the standard normal distribution $N(0,1)$. Equivalently $Z = \frac{X}{\sqrt{Y/n}}$, where $Y \sim \chi^2(\nu)$.

Normal Distribution Theory

t-Distribution

- The *t-distribution* with *n* degrees of freedom (or Student-t), is the distribution of the random variable: $Z = \dfrac{X}{\sqrt{\left(X_1^2 + \cdots X_n^2\right)/n}}$

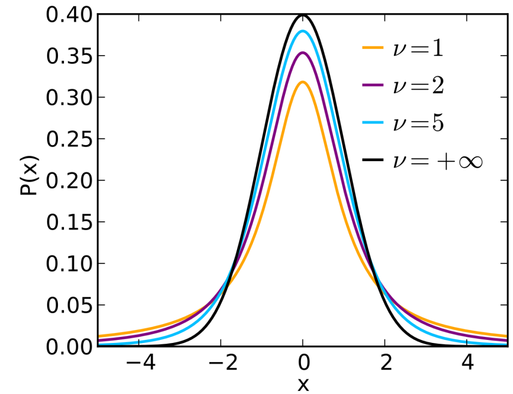
where $X_1, X_2, \ldots, X_n$ are i.i.d., each with the standard normal distribution $N(0,1)$. Equivalently $Z = \dfrac{X}{\sqrt{Y/n}}$, where $Y \sim \chi^2(\nu)$.

Applications: small sample sizes (n≲ 30), and/or limited information

PDF: $f(t) = \dfrac{\Gamma(\frac{\nu+1}{2})}{\sqrt{2\pi}\Gamma(\frac{\nu}{2})}\left(1 + \dfrac{t^2}{\nu}\right)^{-(\nu+1)/2}, -\infty < t < +\infty,$

$$\text{where,} \quad \mu = 0, \sigma^2 = \frac{\nu}{\nu - 2}, \nu > 2$$

t-Distribution



Examples

1. If T follows a t_4 distribution, find t_0 such that: $P(|T|<t_0)=0.90$.

Solution: df = 4, and $|T|<t_0 \Rightarrow -t_0<T<t_0$

(Area = 90%) $\Rightarrow t_0 = 2.132$

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869

t-Distribution

Examples

2. If T follows a t_2 distribution, find t_0 such that: $P(T<t_0)=0.80$.

Solution: df = 2, and $T<t_0 \rightarrow$ Area = 80%

$\rightarrow t_0=1.061$

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869

t-Distribution

Examples

3. Determine the t-critical value such that the lower-tail area =0.025 for a t_3 distribution.

Solution: df = 3, and $T<t_0 \rightarrow$ Area = 2.5%

$\rightarrow t_0=-3.182$

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869

Normal Distribution Theory

F-Distribution

- The F distribution with ν_1 and ν_2 degrees of freedom (or F_{ν_1,ν_2}) is the distribution of the random variable

$$Z = \frac{\left(X_1^2 + \cdots X_{\nu_1}^2\right)/\nu_1}{\left(Y_1^2 + \cdots Y_{\nu_2}^2\right)/\nu_2}$$

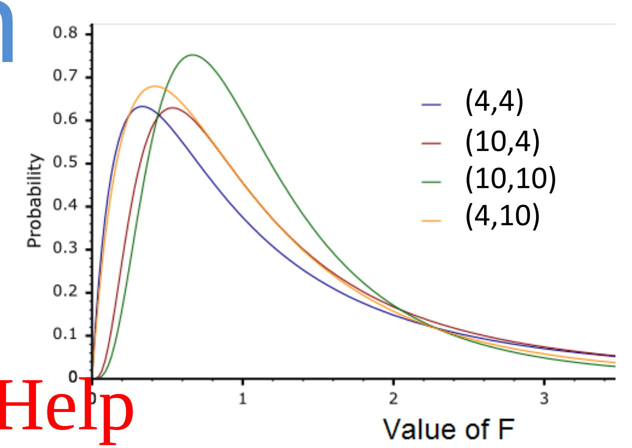
where $X_1, \ldots, X\nu_1, Y_1, \ldots, Y_{\nu_2}$ are i.i.d., each with the standard normal distribution (i.e.,, $Z = \frac{X/\nu_1}{Y/\nu_2}$, where $X \sim \chi^2(\nu_1)$, $Y \sim \chi^2(\nu_2)$.)

- Applications: population variance comparison (ANOVA), likelihood ratio tests, comparing statistical models, etc.

- PDF: $f(x) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}x^{(\nu_1/2)-1}(\nu_2+\nu_1 x)^{-(\nu_1+\nu_2)/2}, x \geq 0$

  where, $\mu = \frac{\nu_2}{\nu_2 - 2}$

F-Distribution



$$\boxed{F_{\alpha,\nu_1,\nu_2}}$$

F Distribution Table (Percentiles)

$1\text{-}\alpha$ $\quad v_2 \quad v_1$

$1-\alpha$	v_2	1	2	3	4	5	6	7	8	9	10
0.95	1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
0.975		647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.7
0.99		4052.2	4999.5	5403.4	5624.6	5763.7	5859.0	5928.4	5981.07	6022.4	6055.9
0.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
0.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
0.99		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
0.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
0.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
0.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23

F-Distribution

Examples

1. Determine the value of $F_{0.05,\,3,\,2}$

Solution: α = 0.05, 1- α =0.95, $v_1 = 3_,\ v_2=2$

→ $F_{0.05,\,3,\,2} = 19.16$

		1	2	3	4	5	6	7	8	9	10
0.95	1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
0.975		647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.7
0.99		4052.2	4999.5	5403.4	5624.6	5763.7	5859.0	5928.4	5981.07	6022.4	6055.9
0.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
0.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
0.99		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
0.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
0.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
0.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23

F-Distribution

Examples

$$F_{1-\alpha,\nu_1,\nu_2} = \frac{1}{F_{\alpha,\nu_2,\nu_1}}$$

2. Determine the value of $F_{0.95,\,3,\,2}$

Solution: α = 0.95, $v_1 = 3$, $v_2 = 2$

→ $F_{0.95,\,3,\,2} = 1/F_{0.05,\,2,\,3} = 1/9.55 = 0.1047$

		1	2	3	4	5	6	7	8	9	10
0.95	1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
0.975		647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.7
0.99		4052.2	4999.5	5403.4	5624.6	5763.7	5859.0	5928.4	5981.07	6022.4	6055.9
0.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
0.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
0.99		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
0.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
0.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
0.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23

F-Distribution

$$F_{1-\alpha,\nu_1,\nu_2} = \frac{1}{F_{\alpha,\nu_2,\nu_1}}$$

Examples

3. Given that $X \sim F_{3,2}$, find P(X< 0.0623)

Solution: α = ?, $v_1 = 3$, $v_2=2$, $1/0.0623 = 16.04$

→ $P(X< 0.0623) = P(1/X > 1/0.0623) = P(1/X > 16.04) = 0.025$

		1	2	3	4	5	6	7	8	9	10
0.95	1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
0.975		647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.7
0.99		4052.2	4999.5	5403.4	5624.6	5763.7	5859.0	5928.4	5981.07	6022.4	6055.9
0.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
0.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
0.99		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
0.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
0.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
0.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23