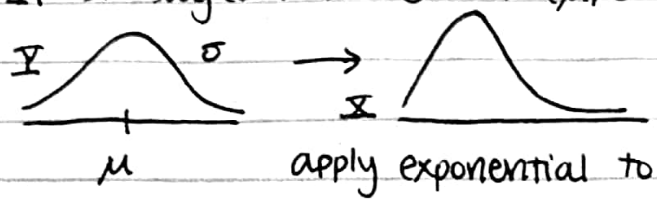
08/26/19

Lecture 13 AMS 131

Lognormal Distribution "Exponential-Normal"

If $Y = \log(X)$ $X > 0$ $\sim N(\mu, \sigma^2)$ then $X \sim \text{Lognormal}(\mu, \sigma^2)$



$X = e^{Y} = e^{N(\mu, \sigma^2)}$

apply exponential to normal distribution

$\psi_Y(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2) = E[e^{tY}] = E[e^{t\log(X)}] = E[X^t]$

$E[X] = \psi_Y(1) = \exp(\mu + \frac{1}{2}\sigma^2)$

$V[X] = \psi_Y(2) - [\psi_Y(1)]^2 = \exp(2\mu + \sigma^2)[e^{\sigma^2} - 1]$

Stock options case study

1 share of a stock, current price : S_0 (known constant)

assumption: the price u time units in the future will be

$$S_u = S_0 e^{Z_u}, \quad Z_u \sim N(\mu \cdot u, \sigma^2 \cdot u)$$

$$= e^{Z_u + \log(S_0)} \rightarrow [Z_u + \log(S_0)] \sim N[\mu \cdot u + \log(S_0), \sigma^2 \cdot u]$$

so $S_u \sim \text{LogNormal}[\mu \cdot u + \log(S_0), \sigma^2 \cdot u]$

$S_u = S_0 \exp[\mu \cdot u + (\sigma\sqrt{u}) \cdot Z]$ where $Z \sim N(0,1)$

price the option to buy 1 share of this stock at price q at time u

risk-neutral pricing $E[S_u] \triangleq S_0$ ← future ← present

u in years, risk-free interest rate r/year (continuously compounding)

the present value of S_u : $S_u = e^{-ru} \cdot E(S_u)$ $E(S_u) = S_0 \exp[\mu \cdot u + \frac{\sigma^2 u}{2}]$

$S_0 = e^{-ru} S_0 \exp[\mu \cdot u + \frac{\sigma^2 u}{2}] \longrightarrow \mu = r - \frac{\sigma^2}{2}$

AMS 131 Lecture 13 (cont.)

value of the option at time u will be $h(S_u)$

$$h(s) = \begin{cases} \underset{\text{price to sell}}{s} - \underset{\text{price you paid}}{q} & \text{if } s > q \\ 0 & \text{else} \end{cases}$$

with $\mu = r - \frac{\sigma^2}{2}$, $h(S_u) > 0$ iff:

$$z > \frac{\log\left(\frac{q}{S_0}\right) - \left(r - \frac{\sigma^2}{2}\right)u}{\sigma\sqrt{u}} \triangleq c$$

present value integral

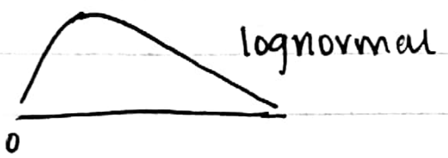
$$e^{-ru} E[h(S_u)] = e^{-ru} \int_c^{\infty} \left[S_0 e^{\left(r - \frac{\sigma^2}{2}\right)u + \sigma z \sqrt{u}} - q\right] \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

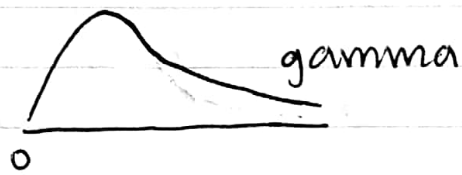
Black-Scholes formula : $S_0 \Phi(\sigma\sqrt{u} - c) - q e^{-ru} \Phi(-c)$

the risk-neutral price of option.

$$c = \frac{\log\left(\frac{q}{S_0}\right) - \left(r - \frac{\sigma^2}{2}\right)u}{\sigma\sqrt{u}}$$

Gamma distribution





$X \sim \Gamma(\alpha, \beta)$ $\quad \beta > 0 \quad$ X continuous on $(0, \infty)$

$$f_X(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \underline{x^{\alpha-1} e^{-\beta x}} \; \underline{I(x>0)} \longleftarrow \text{support}$$

normalizing constant

α: shape parameter (skewness)

β: scale (spread)

gamma function $\Gamma(\alpha) \triangleq \int_0^{\infty} x^{\alpha-1} e^{-x} dx$ no anti derivative in closed form.

↳ continuous generalization of the factorial

(n pos. integer) $\longrightarrow \Gamma(n) = (n-1)!$

$\Gamma(\alpha) \to \infty$ quickly as $\alpha \to \infty$ better to look at log

$$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} = \exp\left[\alpha \ln(\beta) - \ln(\Gamma(\alpha)) + (\alpha-1)\ln(x) - \beta x\right]$$

Stirling's approximation for large x $\quad \Gamma(x) \doteq \sqrt{2\pi}\, x^{x-\frac{1}{2}} e^{-x}$

$$\ln(\Gamma(x)) \doteq \frac{1}{2}\ln(2\pi) + \left(x - \frac{1}{2}\right)\ln(x) - x$$

$X \sim \Gamma(\alpha, \beta)$ ✓ for $t < \beta$

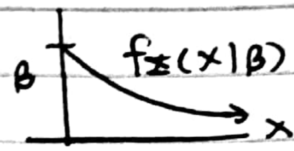
$$\psi_X(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha} \quad E(X) = \frac{\alpha}{\beta} \quad V(X) = \frac{\alpha}{\beta^2} \quad SD(X) = \frac{\sqrt{\alpha}}{\beta}$$

OR $= \left(\frac{\beta}{\beta - t}\right)^{\alpha}$ $\alpha = 1$ $f_X(x|\beta) = \beta e^{-\beta x} I(x > 0)$

Exponential dist. !

$X \sim$ Exponential (β) for $t < \beta$

$$\psi_X(t) = \frac{\beta}{\beta - t} \quad E(X) = \frac{1}{\beta} \quad V(X) = \frac{1}{\beta^2} \quad SD(X) = \frac{1}{\beta}$$



Suppose that arrivals (events) occur according to a Poisson process with rate β per unit time. Set:

Z_K = time until K^{TH} arrival $(K = 1, 2, \ldots)$

$$\left.\begin{aligned} Y_1 &= Z_1 - 0 \\ Y_2 &= Z_2 - Z_1 \\ Y_K &= Y_K - Y_{K-1} \end{aligned}\right\}$$ inter-arrival times: times between arrival events

then, $Y_i \sim$ IID Exponential (β), related to the Geometric with the memoryless property.

↓

$X \sim$ Exponential(β) $t > 0$, $h > 0$

$\rightarrow P(X \geq t + h \mid X \geq t) = P(X \geq h)$

X = time from initial use until a manufactured product fails (ex: lightbulb)

$F_X(x) = P(X \leq x)$ $1 - F_X(x) = P(X > x)$

$= P($"system survives" at least to time $x)$

↳ $1 - F_X(x)$ is the survival function $S_X(x) = 1 - F_X(x)$ in medicine and the reliability function $R_X(x) = 1 - F_X(x)$ in engineering

$F_X(x) = 1 - e^{-\beta x}$ for $x > 0$ $X \sim$ Exponential (β)

$S_X(x) = R_X(x) = e^{-\beta x}$

AMS 131 Lecture 13 (cont.)

instantaneous failure rate = hazard rate

$$H_X(x) = \frac{f_X(x)}{S_X(x)} = R_X(x)$$

this gives $P\left(\text{failure in interval } (x, x+\epsilon) \mid \text{survival to time } x\right)$ small ϵ

"about to fail"

$X \sim$ Exponential (β)

$$H_X(x) = \frac{\beta e^{-\beta x}}{e^{-\beta x}} = \beta \text{ (constant)}$$ the only dist. with constant hazard

If the product has survived until time t, the chance it will survive to time $(t+h)$ is the same as the original chance of surviving from time 0 to h. "the system doesn't remember how long it has survived" (unrealistic)

$X_i \sim$ IID exponential (β) $i = 1, \ldots, n$

then $Y_1 = \min(X_1, \ldots, X_n) \sim$ Exponential $(n\beta)$

Beta distribution $\alpha, \beta > 0$

$$X \sim \text{Beta}(\alpha, \beta) \leftrightarrow f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I(0<x<1)$$

$$B(\alpha, \beta) \triangleq \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$$ beta function $\alpha, \beta > 0$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$ α, β jointly control the shape of the Beta distribution.

$$X \sim \text{Beta}(\alpha, \beta) \quad \psi_X(t) = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!}$$

$$E(X) = \frac{\alpha}{\alpha+\beta} \quad V(X) = \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\beta}{\alpha+\beta}\right)\left(\frac{1}{\alpha+\beta+1}\right)$$

Lecture 13 (cont.) AMS 131

Multinomial distribution (discrete)

population that contains $k \geq 2$ types of elements

ex: Democrat, Republican, Libertarian ...

the proportion of type i is $0 < p_i < 1$ with

$\sum_{i=1}^{k} p_i = 1$ $\quad \underset{\sim}{p} = (p_1, \ldots, p_k)$

take an IID sample of size n from pop.

X_i = # of elements of type i in sample $\quad \sum_{i=1}^{k} X_i = n$ $\quad \underset{\sim}{X} = (X_1, \ldots, X_k)$

PMF $f_{\underset{\sim}{X} | n, \underset{\sim}{p}}(\underset{\sim}{x} | n, \underset{\sim}{p}) = \begin{cases} \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k} & \text{if } \sum_{i=1}^{k} x_i = n \\ 0 & \text{else} \end{cases}$

$\binom{n}{x_1, \ldots, x_k} \triangleq \dfrac{n!}{x_1!\, x_2! \ldots x_k!}$ the multinomial coefficient

$E(X_i) = n p_i$

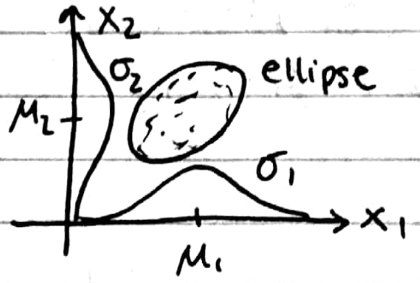
$V(X_i) = n p_i (1 - p_i)$] just like the binomial but...

$C(X_i, X_j) = -n p_i p_j$ negatively correlated $\quad \sum_{i=1}^{k} X_i = n$

When one gets big, the others get small.

Bivariate normal distribution (2D)

$Z_1, Z_2 \sim$ IID $N(0,1)$ with 5 parameters



$-\infty < \mu_1 < \infty$ $\quad 0 < \sigma_1 < \infty$

$-\infty < \mu_2 < \infty$ $\quad 0 < \sigma_2 < \infty$

$-1 < \rho < 1$ correlation

$X_1 = \mu_1 + \sigma_1 Z_1$

$X_2 = \sigma_2 [\rho Z_1 + \sqrt{1-\rho^2}\, Z_2] + \mu_2$

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi \sqrt{1-\rho^2}\, \sigma_1 \sigma_2} \exp\left\{ \frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho \left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 \right] \right\}$$

standard units

AMS 131 Lecture 13 (cont.)

$E(X_1) = \mu_1 \quad V(X_1) = \sigma_1^2$

$E(X_2) = \mu_2 \quad V(X_2) = \sigma_2^2 \qquad \rho(X_1, X_2) = \rho$

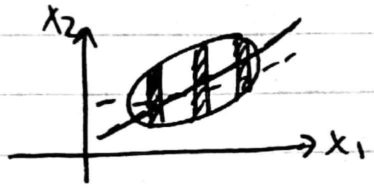
1. $(X_1, X_2) \sim$ Bivariate Normal $\rightarrow$

$\left(\begin{array}{c} X_1, X_2 \\ \text{independent} \end{array}\right) \longleftrightarrow \left(\begin{array}{c} X_1, X_2 \\ \text{uncorrelated} \end{array}\right)$ now this works both directions!

2. $(X_1, X_2) \sim$ Bivariate Normal $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$

$\rightarrow$ conditional distribution of X_2 given $X_1 = x_1$ is univariate normal with mean: $E(X_2 | x_1) = \mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x_1 - \mu_1)$

$V(X_2 | x_1) = (1-\rho^2)\sigma_2^2$



vertical slices $\leftarrow$ also normal

conditional x_2 given x_1

$m = \frac{S_{x_2}}{S_{x_1}}$ (SD line) $\qquad m = r\frac{S_{x_2}}{S_{x_1}} = \hat{\beta}_1$ (regression line)

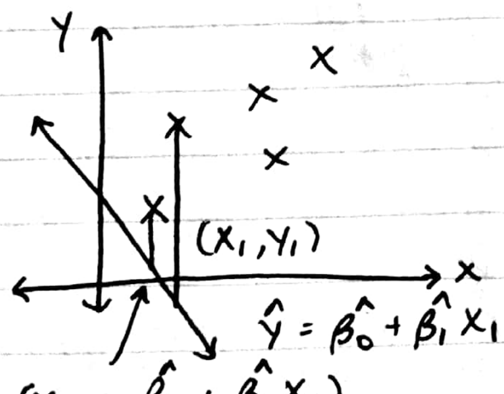
$\hat{X}_2 = \hat{\beta}_0 + \hat{\beta}_1 x_1 \qquad \bar{X}_2 = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_1$

$\hat{\beta}_0 = \bar{X}_2 - \hat{\beta}_1 \bar{X}_1$

$\hat{X}_2 = (\bar{X}_2 - \hat{\beta}_1 \bar{X}_1) + \hat{\beta}_1 X_1$

$= \bar{X}_2 + \hat{\beta}_1 (X_1 - \bar{X}_1) = \bar{X}_2 + r\left(\frac{S_{x_2}}{S_{x_1}}\right)(x_1 - \bar{x}_1)$ sample

$E[X_2 | X_1 = x_1] = \hat{X}_2 = \mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(X_1 - \mu_1)$ theoretical



$(x_1, \hat{\beta}_0 + \hat{\beta}_1 x_1)$

distance: $y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_1)$

$\hookrightarrow \frac{1}{n}\sum_{i=1}^{n} y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$ has pos./neg. values

$g(\hat{\beta}_0, \hat{\beta}_1) = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right]^2$

- find $(\hat{\beta}_0, \hat{\beta}_1)$ to minimize g

$\rightarrow$ least squares line

Lecture 13 (cont.) AMS 131

$\hat{\beta}_1 = r\frac{S_Y}{S_X}$ $\quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ $\quad$ least squares line = regression line

$\beta_1 = \rho\frac{\sigma_2}{\sigma_1}$ $\quad \beta_0 = \mu_2 - \beta_1\mu_1$ $\quad \hat{x}_2 = \beta_0 + \beta_1 x_1$

if we ignore x_1 best prediction is : $(\hat{X}_2)_{\substack{no\\x_1}} = \mu_2 = E(X_2)$

root mean squared (RMSE) $= \sqrt{V(X_2)} = \sigma_2$

now use x_1 to predict x_2 : $(\hat{X}_2)_{\substack{use\\x_1}} = E(X_2 \mid X_1 = x_1) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$

RMSE $= \sqrt{V(X_2 \mid x_1)} = \sigma_2\sqrt{1-\rho^2} \le \sigma_2$ $\quad$ since $-1 < \rho < 1$

RMSE(use x_1) $\le$ RMSE(no x_1)

3. $(X_1, X_2) \sim$ Bivariate Normal $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$

$Y = a_1X_1 + a_2X_2 + b$ $\quad a_1, a_2, b$ constants

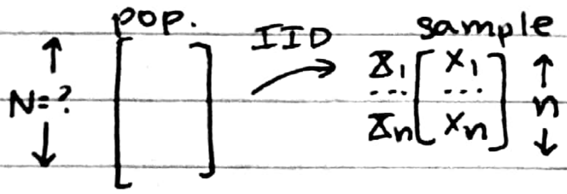
$Y \sim N(a_1\mu_1 + a_2\mu_2 + b,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\sigma_1\sigma_2\rho)$

Large random samples

you draw an IID random sample $X_1, \dots, X_n$ from a population to estimate the pop. mean $\mu = E(X_i)$

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\quad$ intuition, $\bar{X}_n$ is a good estimate of μ

X_i continuous



mean $\mu = E(X_i)$

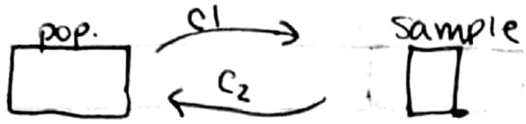
SD $\sigma = \sqrt{V(X_i)}$

AMS 131

Lecture 13 (cont.)

Case 1: pop. known, sample not taken (unknown)

Case 2: sample known, pop. unknown



Most of our example have been case 1

Case 1: probability, Case 2: statistics

general → particular | induction | harder

whole → part | (inference)

deduction (probability) easier

X non-negative R.V. $P(X \geq 0) = 1$

then for all $t > 0$ $\quad P(X \geq t) \leq \dfrac{E(X)}{t}$

- if $E(X)$ is fixed, you can't move more probability into the right tail beyond a point

  $E(X) = 1$, X non negative, $P(X \geq 100) \leq \frac{1}{100}$

  the inequality is sharp, upper bound obtainable

  $E(X) = 1$, put 0.99 prob. on $X = 0$ and 0.01 on $X = 100$

  most of the time it's a crude upper bound

$Y = [X - E(X)]^2 \qquad E[Y] = V[X]$

X R.V with $V(X)$ exists, for every $t \geq 0$

$$P[\,|X - E(X)| \geq t\,] \leq \frac{V(X)}{t^2}$$

X farther from $E(X)$ than t

$E(X) = \mu$
$V(X) = \sigma^2$
$$P\left[\,\left|\frac{X-\mu}{\sigma}\right| \geq 3\right] \leq \frac{1}{3^2} = \frac{1}{9}$$

no more than 11% of probability in any dist. can be more than 3 SDs away from mean.
(normal prob. is 0.3%)

Lecture 13 (cont.) AMS 131

$\bar{X}_n$, $X_i \sim$ IID with some distribution

$E(X_i) = \mu$ $\quad i = 1, 2, \ldots n$ $\quad V(X_i) = \sigma^2 < \infty$

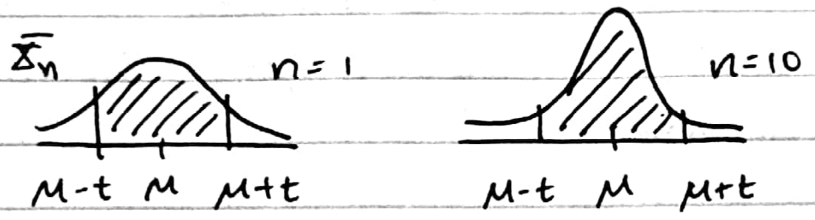
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \underbrace{E(\bar{X}_n) = \mu}_{\text{unbiased}} \quad V(\bar{X}_n) = \frac{\sigma^2}{n}$$

$$P(|\bar{X}_n - \mu| \geq t) \leq \frac{\sigma^2}{nt^2} \quad \text{for all } t > 0$$

$$P(|\bar{X}_n - \mu| < t) \geq 1 - \frac{\sigma^2}{nt^2}$$

for any error tolerance t, as $n \uparrow$, $P(\bar{X}_n$ close to $\mu)$ $\rightarrow 1$, good thing!

less spread, more prob. in fixed region



A sequence of R.Vs $(Z_1, Z_2 \ldots)$ is said to converge in probability to a constant b if for all $\epsilon > 0$

$$\lim_{n \to \infty} P(|Z_n - b| < \epsilon) = 1 \quad \text{denoted } Z_n \xrightarrow{P} b$$

The (weak) law of large numbers : $X_i \sim$ IID, mean μ, variance $\sigma^2 < \infty$

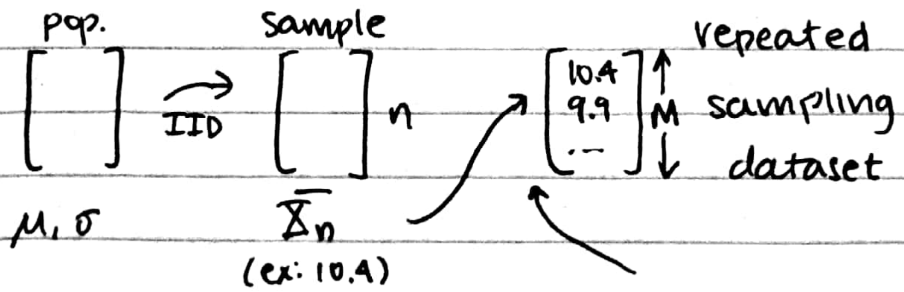
$\bar{X}_n$ is consistent for μ

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \bar{X}_n \xrightarrow{P} \mu$$

Central limit theorem $X_i \sim$ IID $N(\mu, \sigma^2)$ $\quad \sigma < \infty$

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\quad$ mean μ, var: $\frac{\sigma^2}{n}$ $\quad$ normally distributed

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



then take another sample: $\bar{X}_n = 9.9$, repeat alot

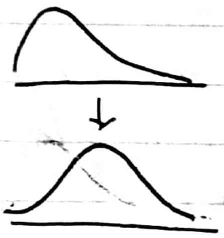
AMS 131 Lecture 13 (cont.)

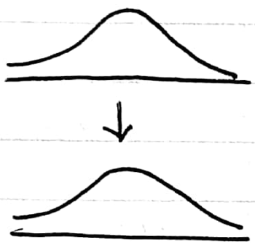
$M \to \infty$, dataset has all $\bar{X}_n$ — say $\sigma = 2$, $n = 25$

long term mean by weak law of large numbers — $E(\bar{X}_n) = \mu = 10$ (example)

long term SD WLLN — $SE(\bar{X}_n) = \sqrt{V(\bar{X}_n)} = \frac{\sigma}{\sqrt{n}} = 0.4$

long term PDF





as long as n is large, $\sigma < \infty$, CLT says in repeated sampling, you will get a normal.

Works for ANY distribution

$X_1, X_2 \ldots$ a sequence of R.V. let F_n be the CDF of X_n → if there exists an F^* CDF such that

$\lim_{n\to\infty} F_n(x) = F^*(x)$ for all x for which $F^*(x)$ is continuous then,

$\bar{X}_n \xrightarrow{D} F^*$ — X_n converges in distribution to F^*

$X_i \sim$ IID any dist. μ mean, $0 < \sigma^2 < \infty$ — $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

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