Lesson 16
Limit Theorems
The Laws of Large Numbers (LLNs)

LLNs provide a framework for an intuitive interpretation of the expectation of a r.v. as the "average value" of the r.v.

LLNs state that the sample average of a large number of i.i.d random variables converges to their expected value.

Weak Law of Large Numbers: Convergence in Probability

Strong Law of Large Numbers: Convergence almost surely)

First proof of LLNs was carried out by Jacob Bernoulli and was published after his death in 1713.

Theorem: (Weak Law of Large Numbers): Let $X_1, X_2, \ldots$ be i.i.d r.v.'s with finite mean $\mathbb{E}[X]$, and let

$$S_n = \sum_{i=1}^{n} X_i$$

Then:

$$\bar{X}_n = \frac{S_n}{n} \xrightarrow{i.p.} \mathbb{E}[X]$$

Proof: (Finite Variance Assumption)

$\sigma_X^2 < \infty$

Taylor Expansion
(Needed for General Proof)

Theorem (Taylor's Theorem)

Assume that $f: \mathbb{R} \to \mathbb{R}$ is continuous and its first n derivatives exist in a neighborhood of 0.

Then:

$$f(x) = f(0) +$$

$$= \sum_{i=1}^{n} \frac{f^{(i)}(0) x^i}{i!} + O(x^{n+1})$$

where $O(x^{n+1})$ satisfies:

General Proof Using Characteristic Functions:

Strong Law of Large Numbers

Theorem: Assume that $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of i.i.d r.v.s with $\mathbb{E}[|X_i|]<\infty$, then

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathbb{E}[X], \text{ i.e.}$$

$$\mathbb{P}\left(\left\{\omega \,\middle|\, \frac{S_n(\omega)}{n} \to \mathbb{E}[X]\right\}\right)=1$$

where $S_n = \sum_{i=1}^{n} X_i$

Proof: Omitted.

For a fixed $\omega \in \Omega$, $\left\{ \frac{S_n(\omega)}{n} \right\}_{n \in \mathbb{N}}$ is a sequence of real numbers. There are three possibilities regarding the convergence of this sequence:

1. The sequence $\frac{S_n(\omega)}{n}$ doesn't converge as $n \to \infty$
2. $\frac{S_n(\omega)}{n}$ converges to a value other than $\mathbb{E}[X]$ as $n \to \infty$

3. $\frac{S_n(\omega)}{n}$ converges to $\mathbb{E}[X]$

as $n \to \infty$.

[The SLLN asserts that the

set of $\omega \in \Omega$ where the

first and second possibilities

hold has a probability

of $\boxed{}$.

Example: Assume that $\{X_i\}$'s

are iid $N(0,1)$. Consider

the event

$$\{\omega \mid X_1(\omega)=1, X_2(\omega)=1, X_3(\omega)=1, \dots\}$$

This event is equivalent to

$$\left\{\omega \;\middle|\; \frac{S_n(\omega)}{n} = \qquad \right\}$$

This event is possible

to occur. What is the probability of this event?

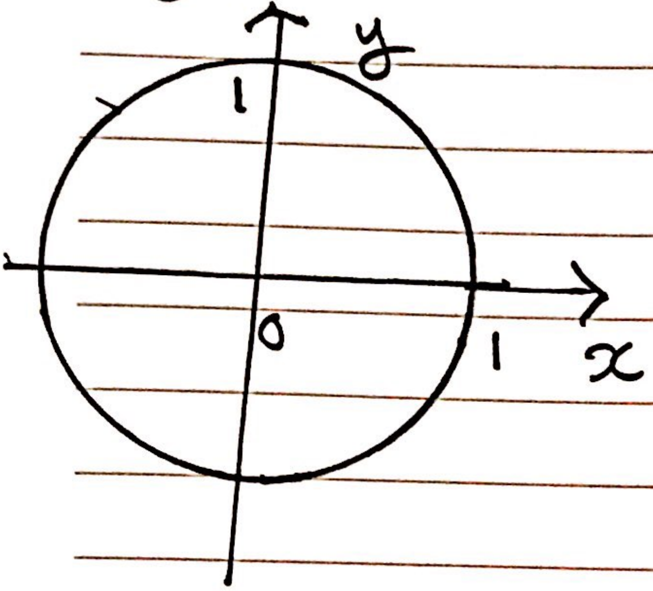
Remark: SSLN guarantees almost sure convergence, but WLLN guarantees convergence in probability. Therefore

$$\text{SLLNs} \xrightarrow{\text{implies}} \text{WLLNs}$$

a.s. i.p.

The LLNs is a basis for the so called Monte-Carlo Simulation Techniques

Example: Monte Carlo Estimation of π



Not hard to prove that

$$\int_0^1 \sqrt{1-x^2}\, dx = \frac{\pi}{4}$$

How to estimate an integral using Monte-Carlo Methods:

$$I = \int_a^b g(x)\,dx = (b-a)\int_a^b \frac{g(x)\,dx}{b-a}$$

Assume $b < a$ and $U \sim \text{Uniform}(a,b)$

$$I = b-a \int_a^b f_U(x)\, g(x)\, dx = (b-a)\, E_U[g(U)]$$

Algorithm:

- Draw an i.i.d $U(a,b)$ sample $U_1, U_2, \dots, U_n$, where n is large.
- Calculate $Y_i = g(U_i)$
- By the LLNs,

$$\bar{Y}_n \xrightarrow[i.p.]{a.s} E_U[g(U)]$$

Therefore, $I \simeq (b-a)\bar{Y}_n$

In particular, for calculating π, the following average should be calculated

$$I = \int_0^1 \sqrt{1-x^2}\, dx = \frac{\pi}{4}$$

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} \sqrt{1-U_i^2}$$

and $\pi \approx \frac{4}{n}\sum_{i=1}^{n}\sqrt{1-U_i^2}$

y,

The Central Limit Theorem

The WLLNs states that

$$\frac{S_n}{n} \xrightarrow{i.p.} \mathbb{E}[X]$$

The SLLNs states that

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathbb{E}[X]$$

Therefore

$$\frac{S_n - n\,\mathbb{E}[X]}{n} \longrightarrow$$

as $n \to \infty$

Therefore, $S_n - n\mathbb{E}[X]$

grows slower than n.

The CLT states that if i.i.d

X_i's have finite non-zero

variance, $S_n - n\mathbb{E}[X]$

scales as $\sqrt{n}$ and the

distribution of $\frac{S_n - n\mathbb{E}[X]}{\sqrt{n}}$

approaches a normal distribution

as $n \longrightarrow \infty$ regardless

of what distributions X_i's have.

i.e.

$$\frac{S_n - n\,\mathbb{E}[X]}{\sigma_X\sqrt{n}} \xrightarrow{d} \mathbb{Z}$$

where $\mathbb{Z} \sim N(0,1)$

Theorem (The Central Limit Theorem):

Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. r.v.'s with finite mean $\mathbb{E}[X]$ and a non-zero finite variance $0<\sigma_X^2<\infty$.

Let $\mathbb{Z}_n = \dfrac{S_n - n\,\mathbb{E}[X]}{\sigma_X\sqrt{n}}$

Then $Z_n \xrightarrow{d} Z$, where

Z is a standard normal r.v.,

i.e.

$$\lim_{n \to \infty} F_{Z_n}(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz$$

$$= \Phi_Z(z) \quad \forall z \in \mathbb{R}$$

Remark: Don't confuse the cdf of standard normal $\Phi_Z(z)$ with its characteristic function $\varphi_Z(\omega)$.

Proof:

Let's see a different form of the CLT that is common in statistics:

$$\frac{S_n - n\,\mathbb{E}[X]}{\sigma_X \sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

Rule of Thumb:

For large n, the sample mean of i.i.d r.v.'s with finite variance $\sigma_X^2 < \infty$ is approximately normal. How large? $n \geq 30$

$$\bar{X}_n \approx \mathcal{N}\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

$$S_n \approx \mathcal{N}(n\mu_X, n\sigma_X^2)$$

Also, show that

$$\sqrt{n}\,(\bar{X}_n - \mu_X) \approx$$

Basically, $\bar{X}$, or the "sample mean" is an "estimate" of the real (population) mean $\mu = \mathbb{E}[X]$. It is a random variable itself. It is a

"good estimate" if it has a small variance.

To make the variance of $\bar{X}$, i.e. $\frac{\sigma_X^2}{n}$ small, we have to increase n, the number of samples.

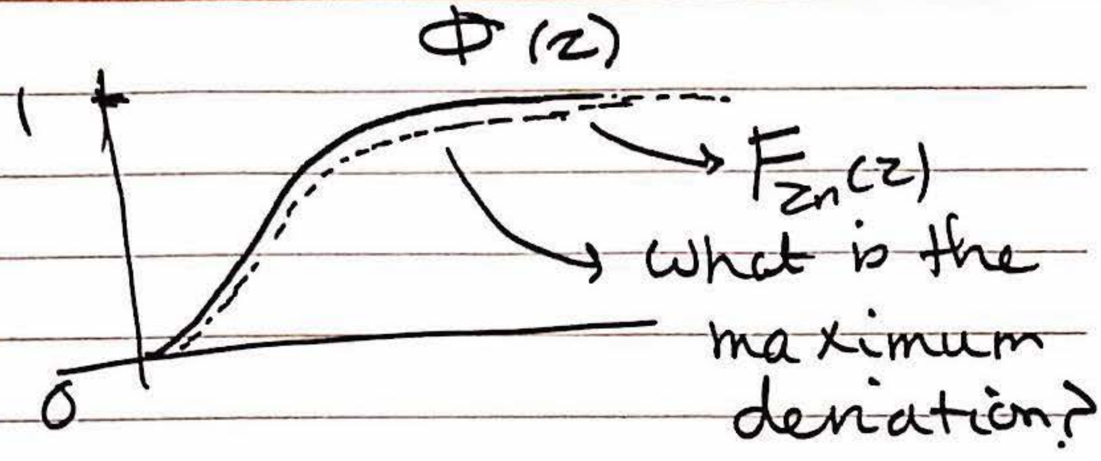
In statistics, $\frac{\sigma_X}{\sqrt{n}}$, the standard deviation of $\overline{X}$, is called the standard error.

Question: How good is the CLT cdf approximation for a given sample size n?

To answer this question, we have to characterize

$$\sup_{z} \left| F_{Z_n}(z) - \Phi(z) \right| \quad \text{as}$$

$n \longrightarrow \infty$.



Theorem (Berry – Esseen):

Given the conditions of the CLT, when the third moment of X is finite, i.e. $\mathbb{E}[|X|^3] < \infty$

$\exists c, \forall z \in \mathbb{R}, \forall n$, s.t.

$$|F_{Z_n}(z) - \Phi_Z(z)| \leq$$

c is known to be in the following interval

$$[.4097, .71]$$

Remark: This means that the maximum error between F_{Z_n} and $\Phi_Z(z)$ has an inverse relationship with $\frac{1}{\sqrt{n}}$.

The Central Limit Theorem for Random Vectors

Definition) Random vectors $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ are independent iff

$$F_{\underline{X}_1 \underline{X}_2 \cdots \underline{X}_n}(\underline{x}_1, \underline{x}_2, \ldots \underline{x}_n) = F_{\underline{X}_1}(\underline{x}_1) F_{\underline{X}_2}(\underline{x}_2) \cdots F_{\underline{X}_n}(\underline{x}_n) \quad \forall \underline{x}_i$$

where the LHS shows the joint distribution of all elements of all $\underline{X}_i$'s, and the RHS involves the joint distribution of the elements of each of $\underline{X}_i$'s, e.g.

$$\underline{X}_1 = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{bmatrix} \qquad F_{\underline{X}_1}(\underline{x}) = F_{Z_1, Z_2, \ldots, Z_m}(x_1, x_2, \ldots, x_m)$$

Remark: A sequence of i.i.d m-vectors $\underline{X}_1, \underline{X}_2, \ldots$ are vectors that are independent and share the same CDF $F_{\underline{X}}$.

Note: $\underline{X}_1, \underline{X}_2, \ldots$ being i.i.d doesn't imply that the members/elements of each vector are independent.

Example: Assume $V_1, V_2, \ldots V_n$ are i.i.d $\mathcal{N}(0,1)$. Also assume $V_i = W_i$ with prob. 1.

Then $\underline{X}_i = \begin{bmatrix} V_i \\ W_i \end{bmatrix}$ are i.i.d.

2-vectors that are independent,

while V_i and W_i are clearly not

independent.

Theorem (CLT for vectors)

Assume that $\underline{X}_i$ are i.i.d

random vectors with the (common) (auto)-

covariance matrix C_X, and

common mean vector $\underline{\mu}_X$.

Then

$$\sqrt{n}\,(\bar{X}_n - \underline{\mu}_X) \xrightarrow{d} \mathcal{N}(0, C_X)$$

Question: How do you define convergence in distribution for random vectors?

Binomial Approximation Using The CLT

Assume that $X_1, X_2, \ldots, X_n$ are i.i.d. Ber(p) r.v.'s. We know that $S_n = \sum_{i=1}^{n} X_i \sim Bin(n,p)$

and $\mathbb{E}[S_n] = np$ and

$Var(S_n) = np(1-p) = npq$

Then for large n

$$S_n \overset{d}{\approx} N(\quad , \quad)$$

So $Bin(n,p) \overset{d}{\approx} N(\quad , \quad)$

Rule of Thumb for The Approximation to Work:

$np \geq 5$ and $nq \geq 5$

Continuity Correction

Assume $X \sim \mathrm{Bin}(n,p)$ and $np \geq 5$ and $npq \geq 5$. We wish to approximate X using Y, $Y \sim \mathcal{N}(np, npq)$.

We know that

$$\mathbb{P}(X \leq k) = \mathbb{P}(X < k+1)$$

where $k = 0, 1, 2, \ldots, n$

When approximating that probability using Y, it is fairly well approximated by $\mathbb{P}(Y \leq K + 1/2)$.

Example: Assume $X \sim Bin(10, 1/2)$.
Calculate $\mathbb{P}(3 \leq X < 6)$.

$n = 10$, $p = 1/2$ $np = 5$, $nq = 5$.

Question: Can we use Poisson Approximation here?

$Y \sim N(5, 5)$.

Continuity Correction

$\mathbb{P}(3 \leq X < 6) \approx \mathbb{P}(\quad Y \quad)$

$= \mathbb{P}\left(\quad \leq \frac{Y - np}{\sqrt{np(1-p)}} \leq \quad \right)$

$= \mathbb{P}(-1.581 \leq Z \leq .3162)$

where $Z \sim N(0,1)$

$= \Phi_Z(0.3162) - \Phi(-1.581)$

≈ 0.5684

Exact: $\mathbb{P}(3 \leq X < 6) = \sum_{x=3}^{5} \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x}$

$= \sum_{x=3}^{5} \binom{10}{x} \left(\frac{1}{2}\right)^{10} = 0.5683$

Difference: 0.0001

Exercise: Assume that

$X_1, X_2, \ldots, X_n$ are i.i.d $\mathcal{N}(\mu, \sigma^2)$.

What is the distribution of

the "sample mean" $\bar{X}_n = \frac{1}{n} \sum X_i$?

Remark: From previous exercises, one can see averaging reduces the variance of i.i.d normals, but doesn't change parameters of Cauchy!

The CLT for

The Chi-Squared Distribution

Recall if $X \sim \text{Gamma}(p, \lambda)$, then

$$f_X(x) = \frac{\lambda^p x^{p-1} e^{-\lambda x}}{\Gamma(p)}$$

when $p = \frac{n}{2}$, $n \in \mathbb{N}$

and $\lambda = 1/2$

Gamma Distribution is called

a χ^2_n (Chi-Square with

n degrees of freedom).

Exercise: Show that if

$X \sim \text{Gamma}(\lambda, p)$, then

$$M_X(s) = \frac{\lambda^p}{(\lambda - s)^p} \qquad s < \lambda$$

Exercise: Show that if $X_1, \ldots, X_n$ are independent and $X_i \sim \text{Gamma}(\lambda, p_i)$ then

$$\sum_{i=1}^{n} X_i \sim \text{Gamma}\left(\lambda, \sum_{i=1}^{n} p_i\right)$$

Corollary: If $X_1, \dots, X_n$ are iid and $X \sim \chi^2_1$, then

$$X_1 + X_2 + \dots + X_n \sim \chi^2_n.$$

The CLT

~~Since~~ χ^2_1 has finite and non zero variance; recall

if $X \sim \text{Gamma}(\lambda, p)$

$$\mathbb{E}[X^k] = \frac{\Gamma(p+k)}{\lambda^k \Gamma(p)}$$

Then: $\mathbb{E}[X] =$

$\mathbb{E}[X^2] =$

$Var(X) =$

Since χ_1^2 is Gamma $(1/2, 1/2)$

then $Var(\chi_1^2) = \quad < \infty$

and $\mathbb{E}[\chi_1^2] =$

Then according to the CLT:

$$\chi^2_n = \sum_{i=1}^{n} X_i \quad , \quad X_i \sim \chi^2_1$$

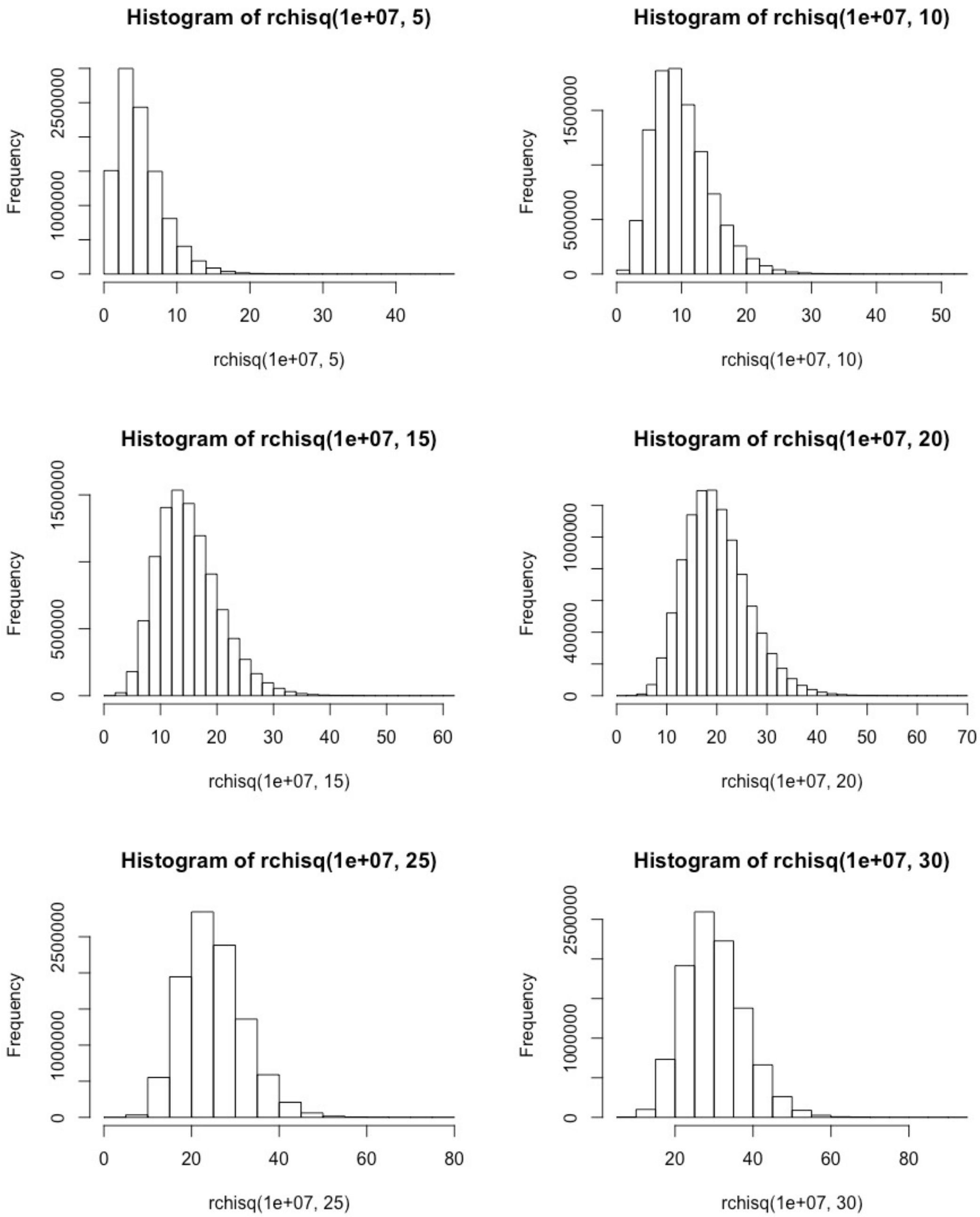
$$\Rightarrow \chi^2_n \overset{d}{\approx} \mathcal{N}\left(n\,\mathbb{E}[\chi^2_1],\ n\,\mathrm{Var}(\chi^2_1)\right)$$

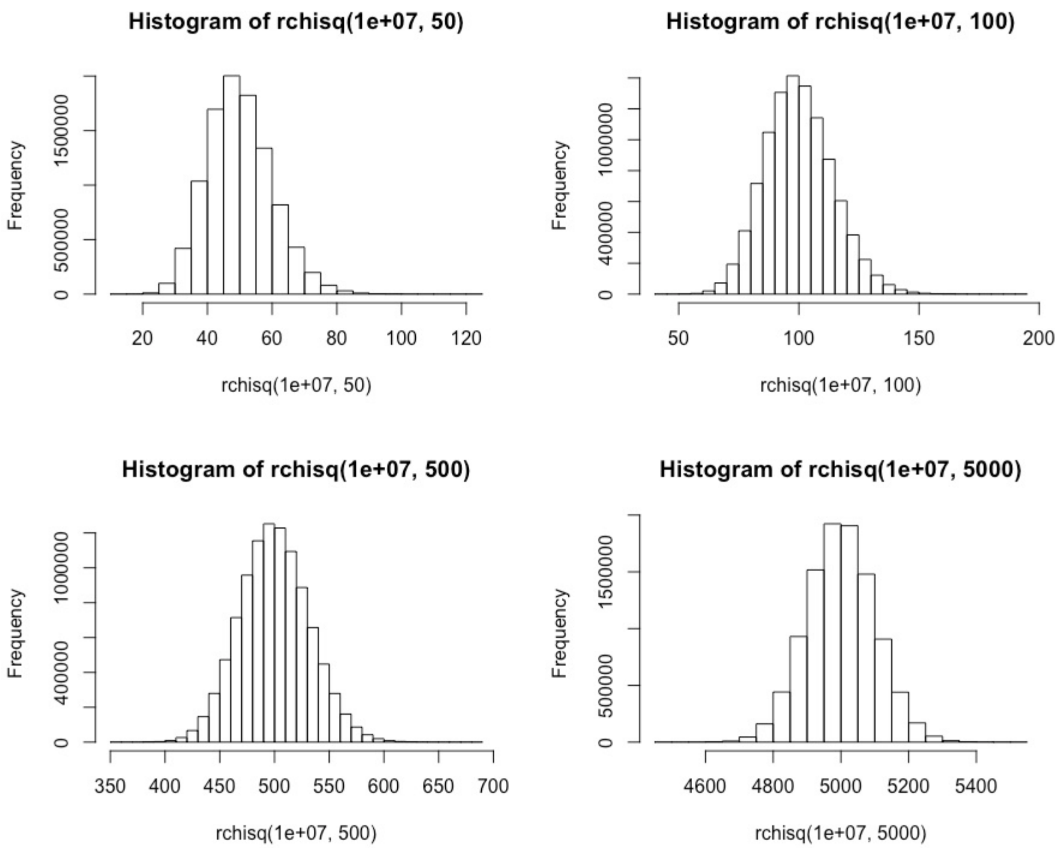
$$\Rightarrow \chi^2_n \overset{d}{\approx} \mathcal{N}(\qquad\qquad)$$

Also $$\frac{1}{n}\chi^2_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\overset{d}{\approx} \mathcal{N}(\qquad)$$

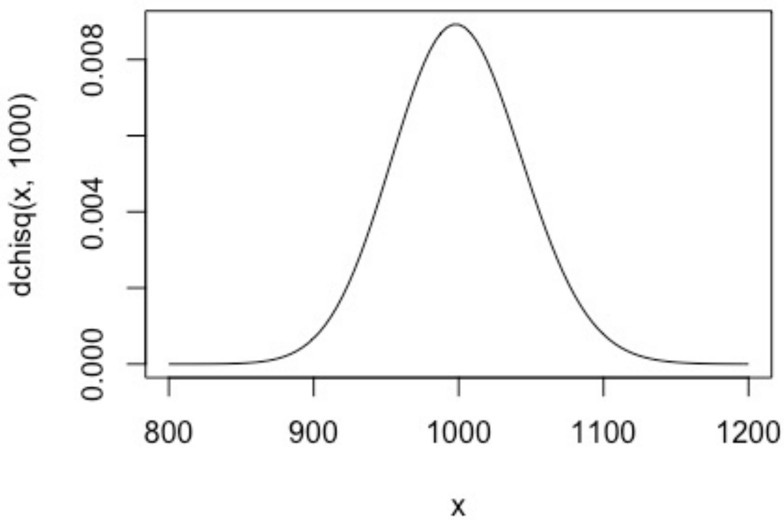
Histograms of 100,000,000 samples from Chi-Square distribution with 5, 10, 15, 20, 25, 30, 50, 100, 500, and 1000 degrees of freedom



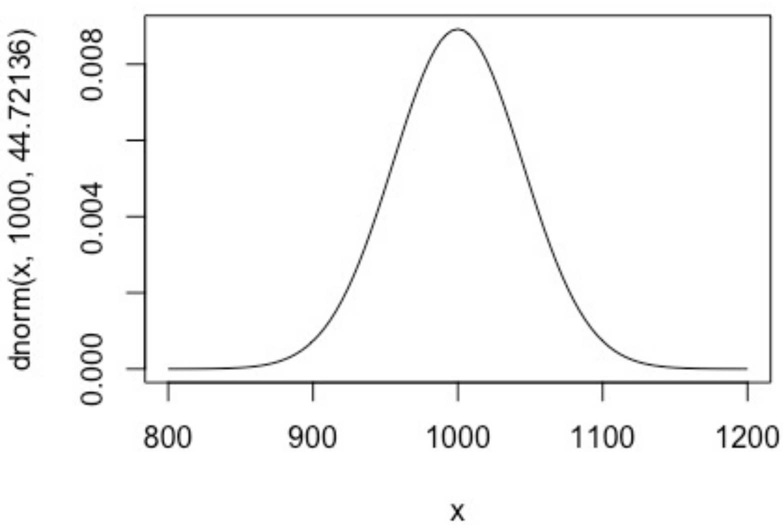



```
> hist(rchisq(10000000,5))
> hist(rchisq(10000000,10))
> hist(rchisq(10000000,15))
> hist(rchisq(10000000,20))
> hist(rchisq(10000000,25))
> hist(rchisq(10000000,30))
> hist(rchisq(10000000,50))
> hist(rchisq(10000000,100))
> hist(rchisq(10000000,500))
> hist(rchisq(10000000,1000))
> hist(rchisq(10000000,5000))
```

Comparison of pdfs of χ^2_{1000} and normal with mean 1000 and variance 2000 (i.e. standard deviation 44.72136)



z



>

```
> x=1:1:200
>curve(dnorm(x,1000,44.72136),800,1200)
> curve(dchisq(x,1000),800,1200)
```