Statistical Theory and Modeling (ST2601)
Lecture 6 - The central theorems, transformations and Monte Carlo

Mattias Villani

Department of Statistics
Stockholm University

mattiasvillani.com @matvil @matvil mattiasvillani

Overview

- **Stochastic convergence**
- **Law of large numbers**
- **Central limit theorem**
- **Transformations of random variables**
- **Monte Carlo simulation**

Stochastic convergence - asymptotics

- Performance of a statistical method in large samples $n \to \infty$.
- Can be a good approximation for finite samples.
- **Sequence of random variables** $X_1, X_2, \ldots, X_n$.
- Example: sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- What happens with X_n as $n \to \infty$?
  - Does it concentrate on a single value?
  - Does the distribution of X_n stabilize?

Convergence in distribution

- The sequence $X_1, X_2, \ldots, X_n$ **converges in distribution** to the random variable X if **"the cdf of X_n starts to look like the cdf of X"** when n gets large.

- $F_n(x)$ is the cdf of X_n
- $F(x)$ is the cdf of X

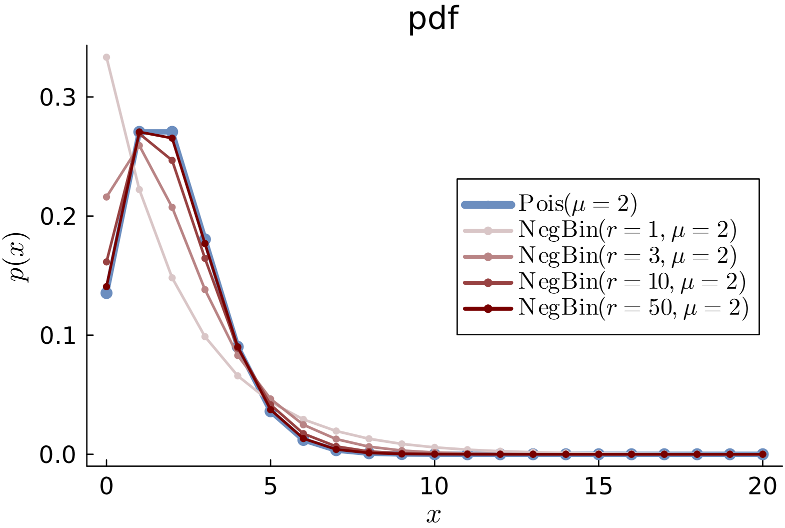
Definition. *A sequence of random variables $X_1, \ldots, X_n$* ***converges in distribution*** *to the random variable X, if*

$$F_n(x) \to F(x) \quad as \quad n \to \infty,$$

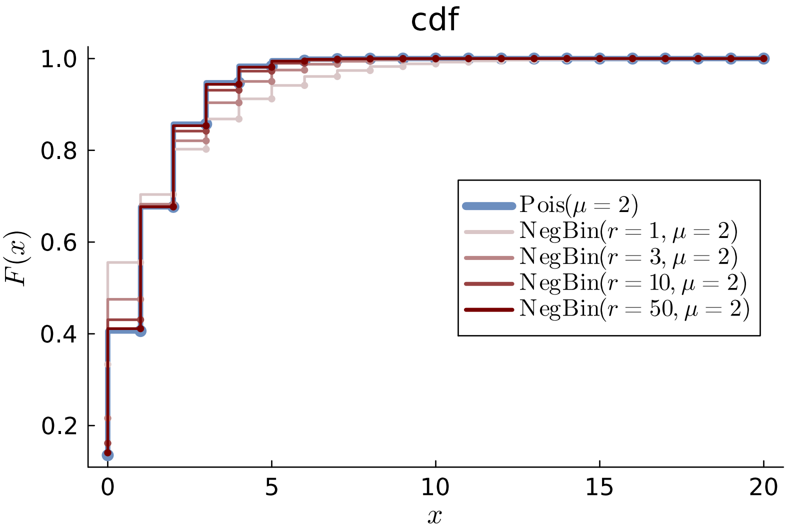
for all x where $F(\cdot)$ is continuous, where $F_n(x)$ and $F(x)$ are the cumulative distribution functions (cdf) of X_n and X, respectively. We then write $X_n \xrightarrow{d} X$.

NegBin converges in distribution to Poisson

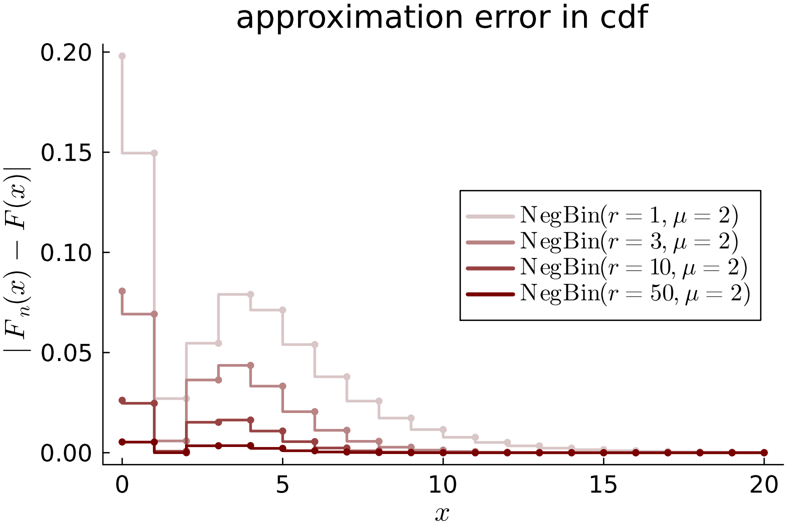
$$\mathrm{NegBin}(r, \mu) \quad \xrightarrow{d} \quad \mathrm{Pois}(\mu) \text{ as } r \to \infty$$



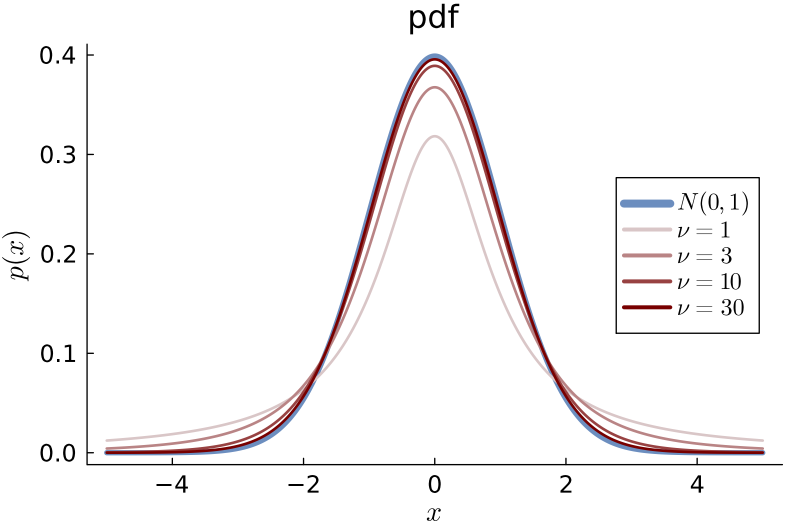
NegBin converges in distribution to Poisson



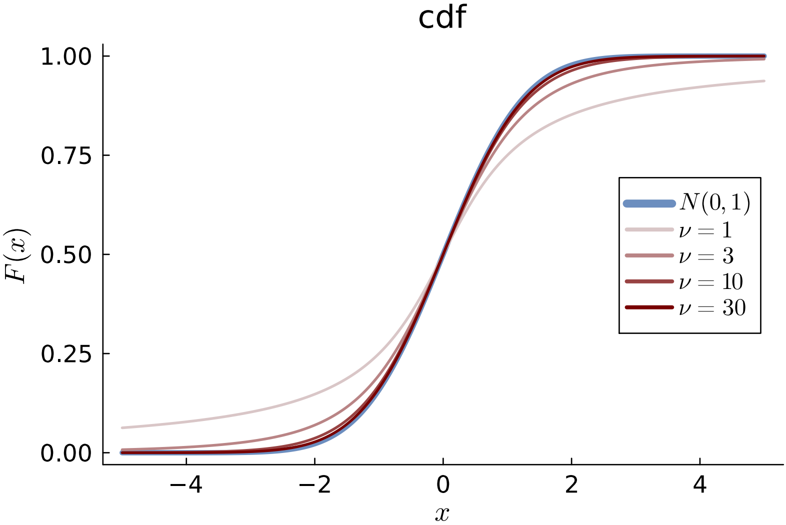
NegBin converges in distribution to Poisson



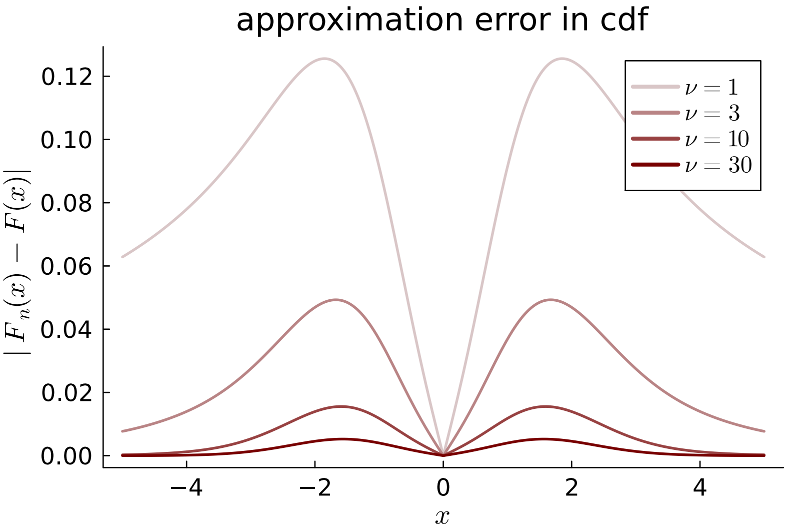
Student-t converges in distribution to $N(0,1)$



Student-t converges in distribution to $N(0, 1)$



Student-t converges in distribution to $N(0,1)$



Limit of a deterministic sequence

- **Mathematical limit** at infinity for **deterministic sequences**

$$\lim_{n\to\infty} x_n = L$$

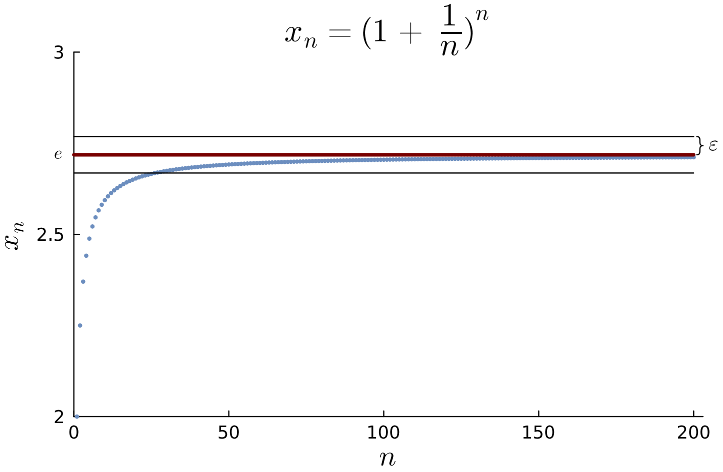
means that we can make sure that

$$|x_n - L| < \epsilon \qquad \Longleftrightarrow \qquad x_n \in (L-\epsilon, L+\epsilon)$$

for any $\epsilon > 0$, by choosing **a large enough** n.

- Example: $x_n = \left(1+\frac{1}{n}\right)^n$, with $\lim_{n\to\infty} x_n = e \approx 2.7183$.
- X_n are random variables, cannot guarantee that $|X_n - L| < \epsilon$.

Limit of a deterministic sequence



Convergence in probability

- The sequence $X_1, X_2, \ldots, X_n$ **converges in probability** to the constant c if **"the distribution of X_n concentrates around c"** when n gets large.

Definition. *A sequence of random variables $X_1, \ldots, X_n$* ***converges in probability to a constant*** *c, if for all $\epsilon > 0$*

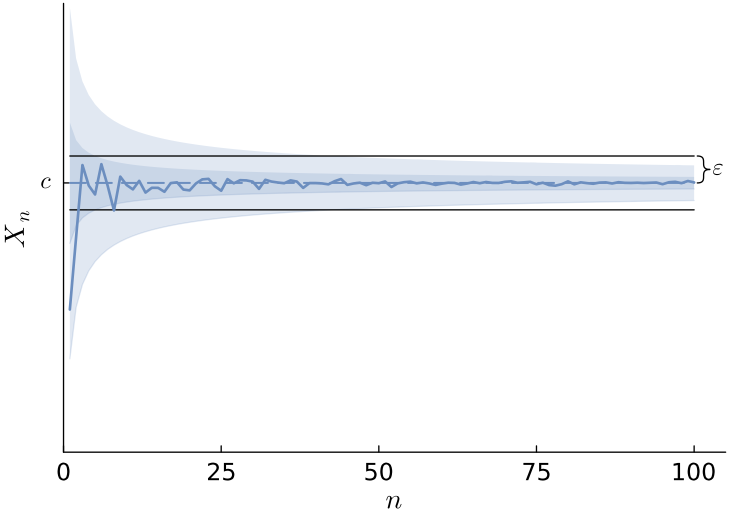
$$\Pr(|X_n - c| > \epsilon) \to 0 \quad \textit{as} \quad n \to \infty.$$

We then write $X_n \xrightarrow{p} c$.

- We can also have convergence in probability to a random variable X instead of a constant; see the prequel book.

Convergence in probability

- 50% and 95% probability intervals.



Law of large numbers

- The **law of large numbers** tells us that the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

converges in probability to the population mean $\mu = \mathbb{E}(X_i)$ as $n \to \infty$.

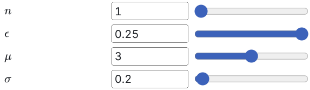
Theorem 4 (law of large numbers).
For independent random variables $X_1, X_2, \ldots$ with finite mean $\mu = \mathbb{E}(X)$ and finite variance we have

$$\bar{X}_n \xrightarrow{p} \mu$$

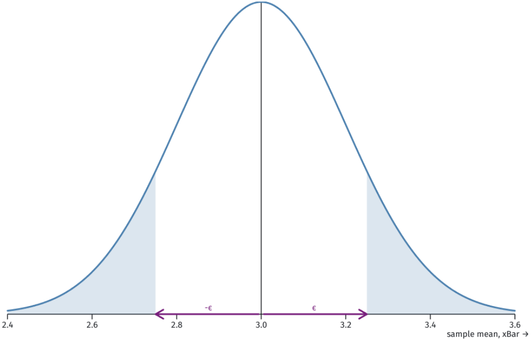
where $\xrightarrow{p}$ denotes convergence in probability, i.e., for all $\epsilon > 0$

$$\Pr(|\bar{X}_n - \mu| \geq \epsilon) \to 0 \quad \text{as} \quad n \to \infty \tag{5.3}$$

Law of large numbers - widget



$$P(|\bar{X}_n - \mu| > 0.25) = 0.2113$$



Central limit theorem

- The **central limit theorem** tells us that the sample mean $\bar{X}_n$ **converges in distribution** to a normal distribution.

> **Theorem 6** (central limit theorem - informal version).
> *Let $X_1, X_2, \ldots$ be iid random variables with finite mean μ and variance σ^2. Then for large n,*
>
> $$\bar{X}_n \overset{\text{approx}}{\sim} \mathrm{N}(\mu, \sigma^2/n)$$

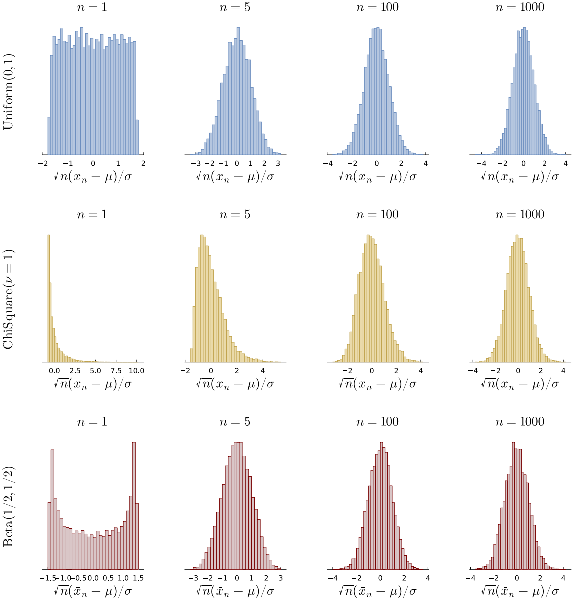
- Have to **standardize** to avoid a degenerate distribution:

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{(\bar{X}_n - \mu)}{\sigma}$$

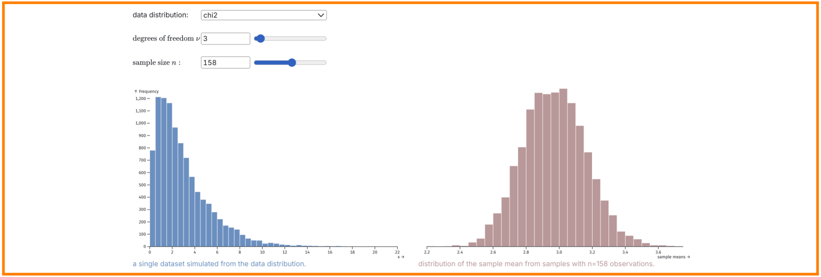
- Formal version

> **Theorem 5** (central limit theorem).
> *Let $X_1, X_2, \ldots$ be iid random variables with finite mean μ and variance σ^2. Then*
>
> $$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathrm{N}(0,1),$$
>
> *as $n \to \infty$, where $\xrightarrow{d}$ denotes convergence in distribution.*

Central limit theorem



Central limit theorem - widget



Transformations of random variables

- **Known**: the distribution of X is $f(x)$
- **Wanted**: the **distribution of a transformed variable**

$$Y = g(X)$$

- Why? We often need to **transform the data**.
- Bayes: we often need to transform parameters.
- Examples:
  - Linear: $Y = a + b \cdot X$
  - Log: $Y = \log(X)$
  - Logit: $Y = \log\left(\frac{X}{1-X}\right)$

Transformations of random variables - example

- Example:
  - **pdf**: $f_X(x) = 3x^2$ for $0 \leq x \leq 1$
  - **cdf**: $F_X(x) = \int_0^x 3t^2 \mathrm{d}t = \left[t^3\right]_0^x = x^3$
- **Linear transformation**: $Y = 2 + 3X$
- **cdf** of Y:

$$F_Y(y) = \Pr(Y \leq y) = \Pr(2 + 3X \leq y) = \Pr\left(X \leq \frac{y-2}{3}\right)$$
$$= F_X\left(\frac{y-2}{3}\right) = \left(\frac{y-2}{3}\right)^3$$

- **pdf** of Y

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X\left(\frac{y-2}{3}\right) = f_X\left(\frac{y-2}{3}\right) \cdot \frac{1}{3}$$
$$= 3\left(\frac{y-2}{3}\right)^2 \cdot \frac{1}{3} = \left(\frac{y-2}{3}\right)^2 \quad \text{for } 2 \leq y \leq 5$$

Transformations of random variables - example

- A little more general: **linear transformation**: $Y = a + bX$
- **pdf** of Y

$$F_Y(y) = \Pr(Y \leq y) = \Pr(a + bX \leq y) = \Pr\left(X \leq \frac{y-a}{b}\right) = F_X\left(\frac{y-a}{b}\right)$$

- **cdf** of Y

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X\left(\frac{y-a}{b}\right) = f_X\left(\frac{y-a}{b}\right) \cdot \frac{1}{b}$$

- We computed the **inverse transformation**, i.e. solved for x

$$y = a + bx \qquad \Longleftrightarrow \qquad x = \frac{y-a}{b}$$

- General: if $g(x)$ is an **invertible function**

$$y = g(x) \qquad \Longleftrightarrow \qquad x = g^{-1}(y)$$

where $g^{-1}(y)$ is the **inverse function**.

Transformations of random variables

Transforming variables - change-of-variable formula

Let $X \sim f_X(x)$ and

$$Y = g(X)$$

an invertible monotonically increasing or decreasing transformation with continuous derivative and inverse transformation

$$X = g^{-1}(Y).$$

The density of Y is then

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right|$$

- If $Y = g(X)$ is piecewise monotone, handle each piece separately and sum up.
- Example 3 on Wikipedia on transformations uses this on:
  - $X \sim N(0, 1)$
  - $Y = X^2$ which is monotone on $(-\infty, 0)$ and $[0, \infty)$
  - Result: $Y \sim \chi^2(\nu = 1)$

Transformations of random variables - example

- Let $X \sim N(\mu, \sigma^2)$ with pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

- Let $Y = \exp(X)$ with inverse transformation $X = \log(Y)$ with derivative

$$\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) = \frac{\mathrm{d}}{\mathrm{d}y} \log(y) = \frac{1}{y}$$

- Then

$$f_Y(y) = f_X\left(\log(y)\right) \cdot \frac{1}{y} = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\log(y)-\mu)^2\right)$$

for $y > 0$.

- We have shown: if $X \sim N(\mu, \sigma^2)$ then

$$\exp(X) \sim \mathrm{LogNormal}(\mu, \sigma^2)$$

Monte Carlo simulation

- Let $X \sim f(x)$.
- Compute $\mathbb{E}\left(g(X)\right)$ for some function $Y = g(X)$ by simulation.
- Key idea: **law of large numbers**.
- Simulate $x_1, \ldots x_N \overset{\text{iid}}{\sim} f(x)$

$$\frac{1}{m}\sum_{i=1}^{m} g(x_i) \xrightarrow{p} \mathbb{E}\left(g(X)\right)$$

- Monte Carlo to **compute tail probability** $\Pr(X > c)$

$$g(x) = \begin{cases} 1 & \text{if } x > c \\ 0 & \text{if } x \leq c \end{cases}$$

$$\begin{aligned} \mathbb{E}\left(g(X)\right) &= \int_{-\infty}^{\infty} g(x) f(x) \mathrm{d}x = \int_{-\infty}^{c} 0 \cdot f(x) \mathrm{d}x + \int_{c}^{\infty} 1 \cdot f(x) \mathrm{d}x \\ &= \int_{c}^{\infty} f(x) \mathrm{d}x = \Pr(X > c) \end{aligned}$$

Monte Carlo accuracy via the CLT

- Central limit theorem (informal)

$$\frac{1}{m}\sum_{i=1}^{m} g(x_i) \overset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma^2}{N}\right)$$

where

$$\mu = \mathbb{E}\left(g(X)\right)$$

and

$$\sigma^2 = \mathbb{V}(g(X))$$