Sampling Distributions and Convergence Ideas

Sujit Sandipan Chaugule[1*], Dr. Amiya Ranjan Bhowmick[2]

[1*]Department of Pharmaceutical Sciences and Technology, Institute of Chemical Technology, Mumbai

[2]Department of Mathematics, Institute of Chemical Technology, Mumbai

Understanding sampling

In a typical statistics classroom, the teacher often begins by discussing either a coin-tossing experiment or by collecting data randomly from a normally distributed population, with the goal of estimating the probability of success or the population mean, respectively. In both cases, no actual statistical sampling occurs; instead, students are asked to imagine a hypothetical experiment carried out by the teacher, calculating the sample proportion or sample mean based on imagined sample data. Consequently, students often miss the actual connection between fixed sample values (obtained once data is collected) and theoretical probability density functions (which are purely mathematical constructs)
This chapter addresses the gap in understanding statistical concepts by emphasizing the importance of real-time random experiments, similar to those conducted in Physics, Chemistry, Biology, or Engineering courses. By leveraging software capable of simulating various probability distributions, statistical comprehension can be significantly enhanced.To facilitate an understanding of randomness in computations based on sample data, this chapter utilizes computer simulations. Concepts that are often grasped only intuitively are reinforced through simulated experiments, transforming abstract ideas into concrete understanding. The Julia codes presented in this document were developed while attending the Statistical Computing course as part of the Multidisciplinary Minor Degree Programme in Machine Learning and Artificial Intelligence, offered by the Department of Mathematics, Institute of Chemical Technology, Mumbai, under the framework of the National Education Policy (NEP) 2020. These implementations illustrate theoretical statistical concepts through practical simulation, providing a computational approach to statistical learning.

Writing the first computer simulation

If we set up a simulation experiment, the first task is to assume the population distribution. In this case, we start our discussion with the normal distribution and we use the symbol $f(\cdot)$ to represent the population PDF or PMF throughout this chapter. Suppose that we consider problem of estimating the mean of the population distribution $\mathcal{N}(\mu, \sigma^2)$.

In [1]:
```
using Plots, Statistics, StatsBase, StatsModels, Distributions
using LaTeXStrings
```

In [2]:
```
mu = 3  # population mean
sigma = 1 # population standard deviation

f(x) = pdf.(Normal(mu, sigma), x)  # define the function

# by alternate we can define the functionloc
```

```
function f(x)
    pdf.(Normal(mu, sigma), x)
end

plot(f, -3, 9, color = "red", lw = 2, xlabel = "x",
     ylabel = "f(x)", label = "")
```

Out[2]:

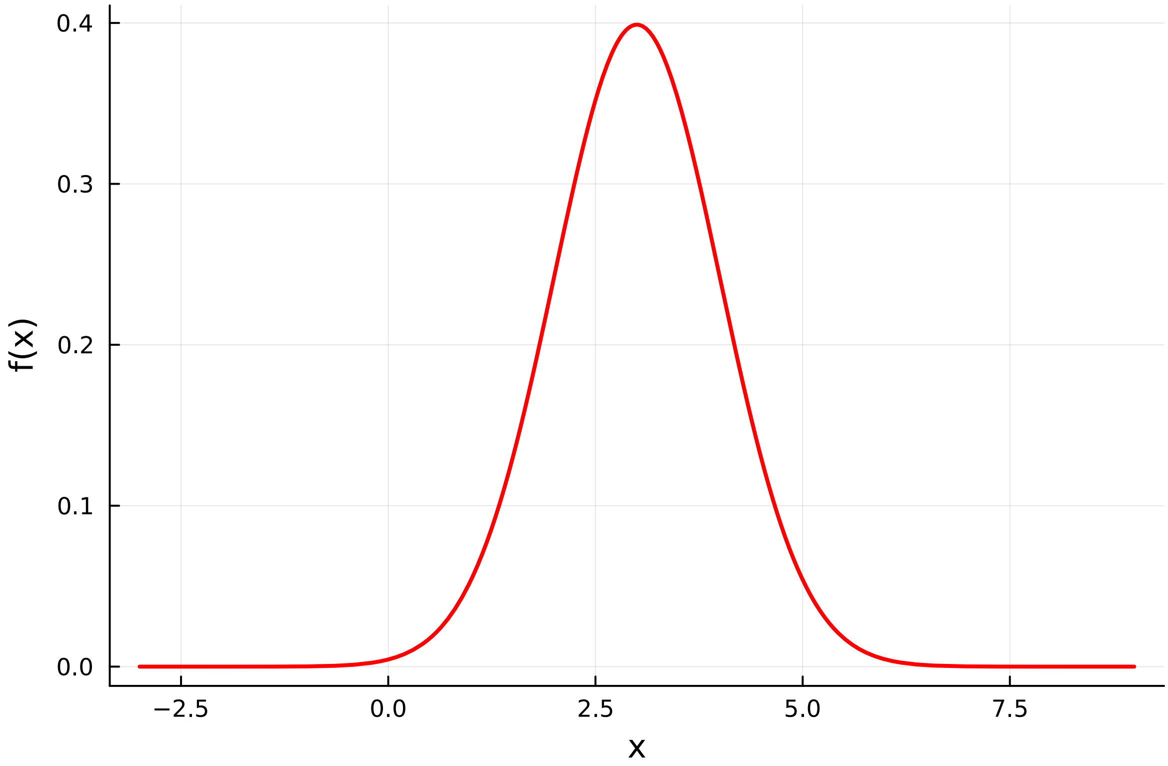

Figure 1: The population probability density function. For simulation study, we consider the population distribution characterized by the normal distribution with mean $\mu = 3$ and variance $\sigma^2 = 1$

Step- I

- Fix the sample size n
- Draw a random sample of size n from the population PDF
- Draw the sample multiple times for different choices of n and also draw the histogram for each sample data.
- Overlay the population density function on the histograms to show that the histograms are actually acting as a good approximation of the theoretical PDF.

In practice while teaching, the drawing of the histogram of the sampled data has been a very effective way to bridge the connection between the theoretical PDF and the sample.

In [3]:

```
n = 50  # sample size
x = rand(Normal(mu,sigma), n)
print(x)
```

```
[3.279546645884055, 0.23868704320889078, 5.038120933270651, 3.869052410116791, 4.58442
6384827575, 3.325088471467067, 4.291955522955943, 2.180883402574202, 2.100034622173720
5, 1.750282569298923, 3.232588358836467, 3.4280391051524166, 2.3704381925814584, 2.587
5812958929445, 3.7042858289993554, 1.6798146992255334, 3.750809645732336, 3.0551165871
42482, 2.4449203133945066, 3.3561730796348597, 3.3953476290726665, 3.202378208419005,
1.835083267599669, 3.167085990162315, 1.9507889210098388, 3.236554576070661, 2.5482346
07354535, 3.8203449298508305, 4.182899245914994, 2.17949430372369, 1.7997567687320761,
2.6360600500862312, 2.682621635861853, 3.681518053971609, 3.1840854048388274, 3.923496
6401720075, 1.5772605140389855, 3.2096690662229554, 3.738203128253558, 3.8979138939148
64, 3.201879986487523, 2.985927832694172, 3.0469610423443143, 2.896870307119677, 4.001
518600563768, 1.6321280804678597, 2.3159968952525722, 3.9216307432238438, 2.7401973167
648097, 3.0882881758942977]
```

In [4]:
```julia
histogram(x, normalize = true , xlabel = "x", ylabel = "density",
    title = "Histogram of x", label = "")
plot!(x->f(x), color = "red", lw = 2, label = "" )
```

Out[4]:

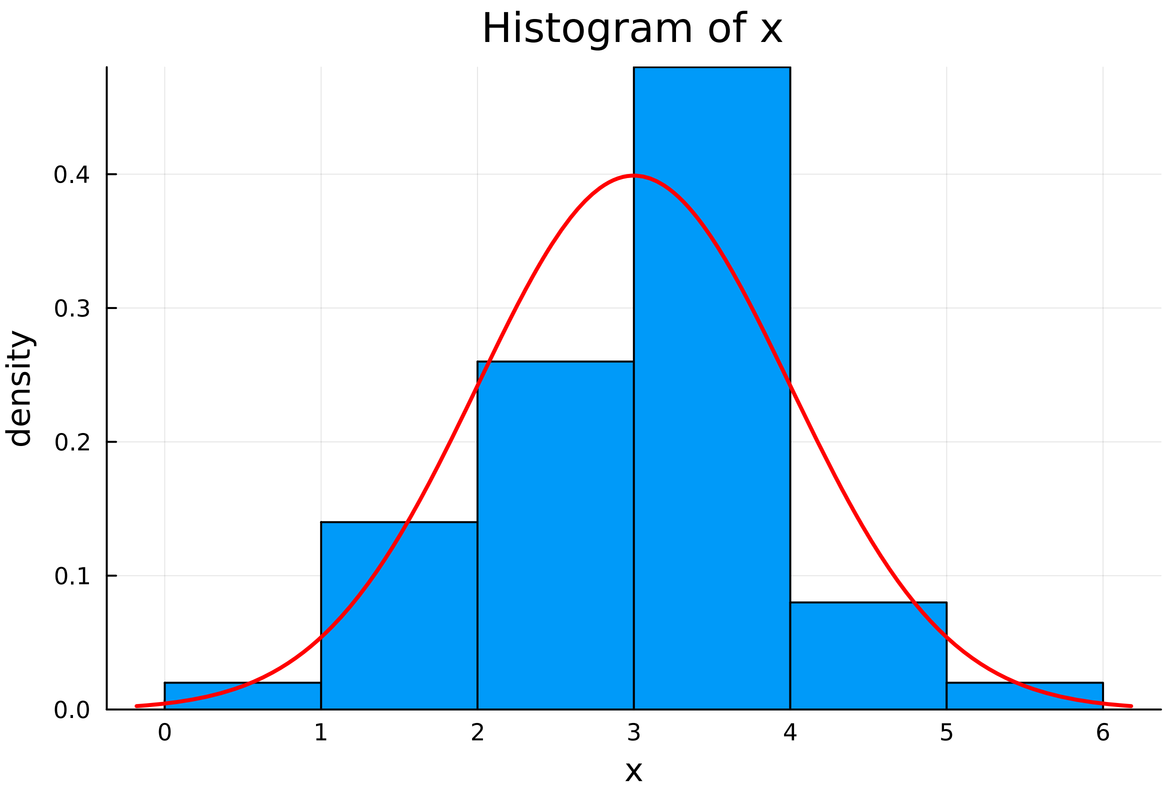

Figure 2:The histogram of the simulated data is displayed, with the population PDF $\mathcal{N}(3,1)$ overlaid. Users are encouraged to execute the following Julia code for different sample sizes and observe the varying shapes of the histograms. Additionally, the effect of the `bins =` option in the `histogram` function can be explored to analyze its impact on the histogram's appearance.

Step - II

Compute Sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and also sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

In [5]:
```julia
sample_mean = mean(x)
sample_variance = var(x)
println("The sample mean is: $sample_mean")
println("The sample variance is: $sample_variance")
```

```
The sample mean is: 2.998960818569084
The sample variance is: 0.8252602910919393
```

Step - III

Invariably, the sample mean will keep on changing as we execute the step- I and step- II multiple times as different run will give different set of random samples. To understand the sampling distribution of the sample mean, we repeat the above process $m = 1000$ times and approximate the actual probability density of the sample mean by using histograms.

In [6]:
```
m = 1000
sample_mean = zeros(m)

for i in 1:m
    x = rand(Normal(mu, sigma), m)
    sample_mean[i] = mean(x)
end

histogram(sample_mean, normalize = true, xlabel = L"\bar{X_n}",
          ylabel = "density", title = "n = $n", label = "")
scatter!([mu],[0], color = "red", markersize = 14, label = "")
scatter!([mean(sample_mean)],[0], color = "blue", markersize = 8,
    label = "")
```

Out[6]:

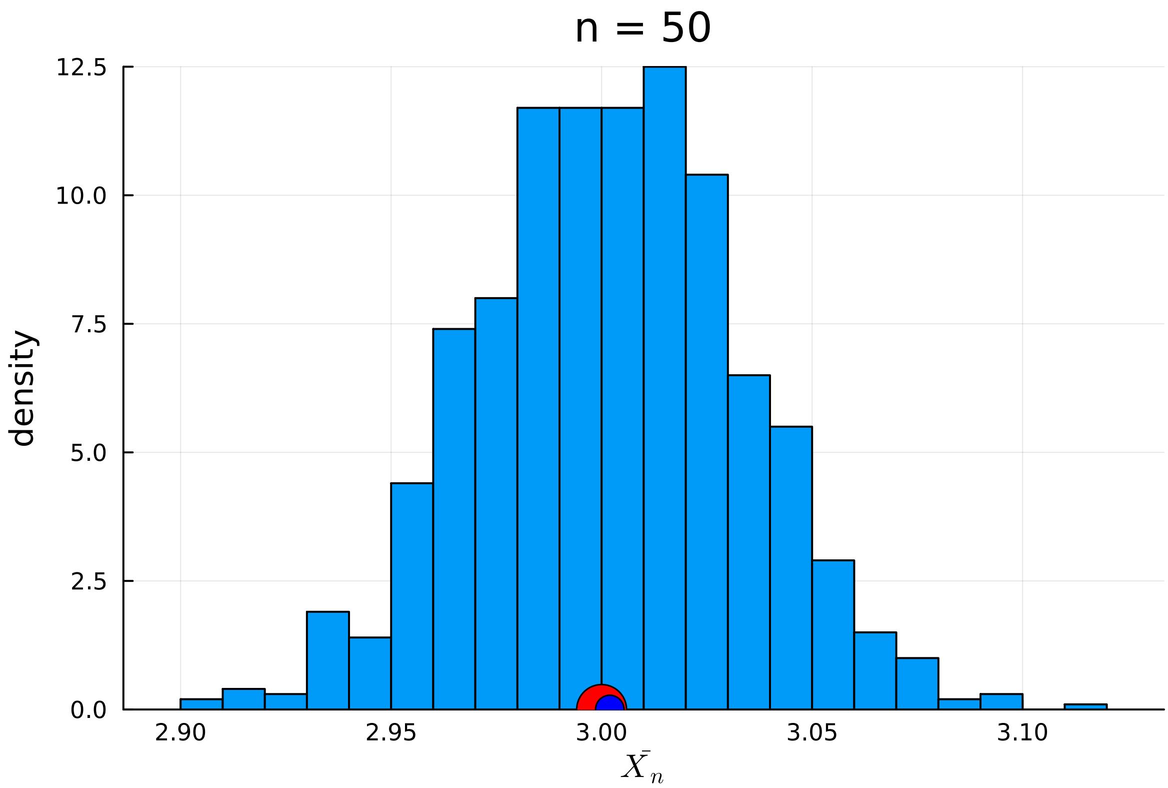

Figure 3: The sampling distribution of $\overline{X}_n$ approximated by histogram based on $m = 1000$ replications. The population distribution is governed by $\mathcal{N}(3, 1)$. Students are encouraged to modify the variables m and n and see the resulting histograms. Check that all the histograms look bell shaped.

In [7]:
```
println("The average of the 1000 many sample mean values is: ",
        mean(sample_mean))
```

```
The average of the 1000 many sample mean values is: 3.001942193285571
```

Sampling distribution of $\overline{X}_n$

> In the above simulation experiment, what are your observations if you perform the following:
>
> - Fix m, how the shapes of histogram change as n increases. (Keep an eye on the *x*-axis of the histogram)
> - Fix n. How the shapes of the histogram changes if m is small and (m) is large.
> - Does the bell-shaped pattern change if you change μ and σ?

Step - IV

Let us do this experiment for different sample sizes and see how the distribution behaves. Our theory suggested that $\mathbb{E}(\overline{X_n}) = \mu$, that is, the sample mean is an unbiased estimator of population mean and the expected value does not depend on n.

In [8]:
```
plt = plot(layout=(1, 3), size=(800, 400))
n_vals = [3,10,25]
m = 1000
for (idx, n) in enumerate(n_vals)
    sample_mean = zeros(m)
    for i in 1:m
        x = rand(Normal(mu, sigma), n)
        sample_mean[i] = mean(x)
    end
    histogram!(plt, sample_mean, normalize = true, xlabel = "sample mean",
     ylabel = "density", title = "n = $n", label = "", subplot = idx)
    scatter!([mu],[0], color = "red", markersize = 14, label = "",
    subplot = idx)
    scatter!([mean(sample_mean)],[0], color = "blue", markersize = 8,
    label = "", subplot = idx)
end

display(plt)
```

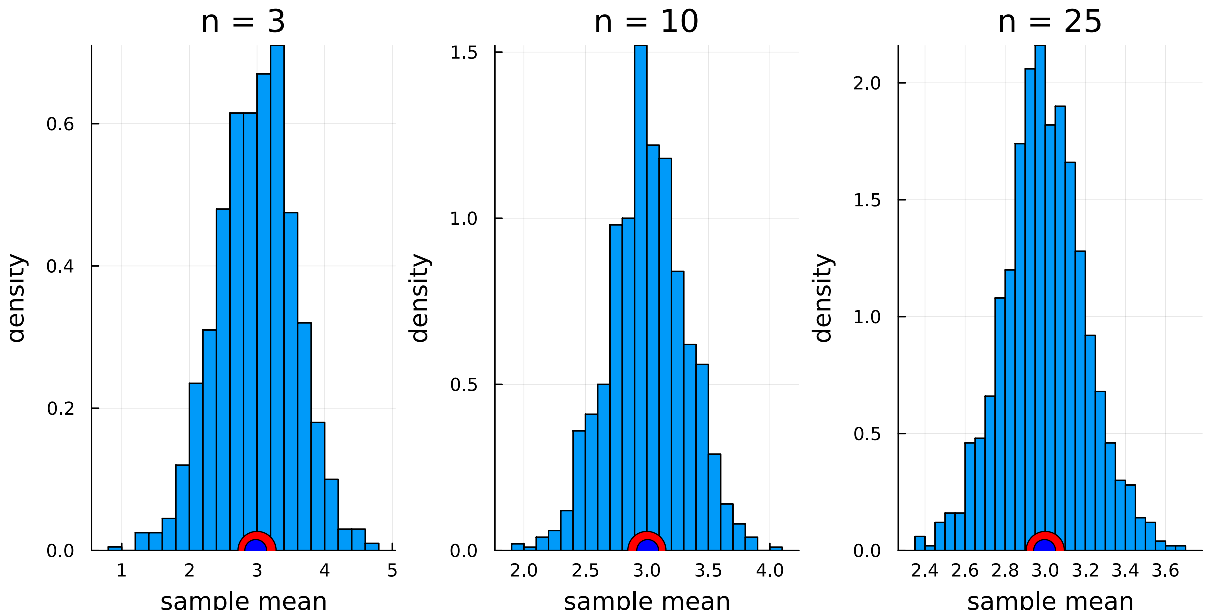

Figure 4: The sampling distribution of the sample mean for different choices of n. The population distribution is considered as the normal distribution with mean 3 and variance 1. It is important to note the average of the

simulated sample mean values (blue color dot) coincides with the true mean of the population (marked as red colored dot).

> **Expectation is equivalent to averaging using simulation**

In the simulation (Step - IV), the true population mean is indicated using the red dot and the average of the sample means based on 1000 replications is shown using a blue circle. The averaging of the sample mean values can be thought of as an expected value of $\overline{X_n}$, that is, $\mathbb{E}(\overline{X_n})$.This simulation gives an idea that $\mathbb{E}(\overline{X_n})$ is equal to the true population mean μ (here it is 1). Although not a mathematical proof, such simulation outcomes give an indication of the unbiasedness of the sample mean.

Step - V

In theory, we have computed by using Moment Generating Functions that

$$\overline{X_n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Let us check whether the theoretical claim can be verified by using simulation.

In [9]:
```
using Plots, Distributions

mu = 3      # Mean of population distribution
sigma = 1   # Standard deviation
n_vals = [3, 10, 25, 50, 100, 250]  # Different sample sizes
m = 1000    # Number of simulations

plt = plot(layout=(2, 3), size=(800, 600))

for (idx, n) in enumerate(n_vals)
    sample_mean = zeros(m)
    for i in 1:m
        x = rand(Normal(mu, sigma), n)
        sample_mean[i] = mean(x)
    end
    histogram!(plt, sample_mean, normalize=:pdf, xlabel="Sample Mean",
        ylabel="Density", title="n = $n", label="", subplot=idx)
    scatter!([mu], [0], color="red", markersize=14, label="", subplot=idx)
    scatter!([mean(sample_mean)], [0], color="blue", markersize=8, label="",
        subplot=idx)
    x_vals = range(minimum(sample_mean), maximum(sample_mean),
        length=1000)
    pdf_normal = pdf.(Normal(mu, sigma / sqrt(n)), x_vals)
    plot!(x_vals, pdf_normal, color="red", lw=2, label="", subplot=idx)
end

display(plt)
```

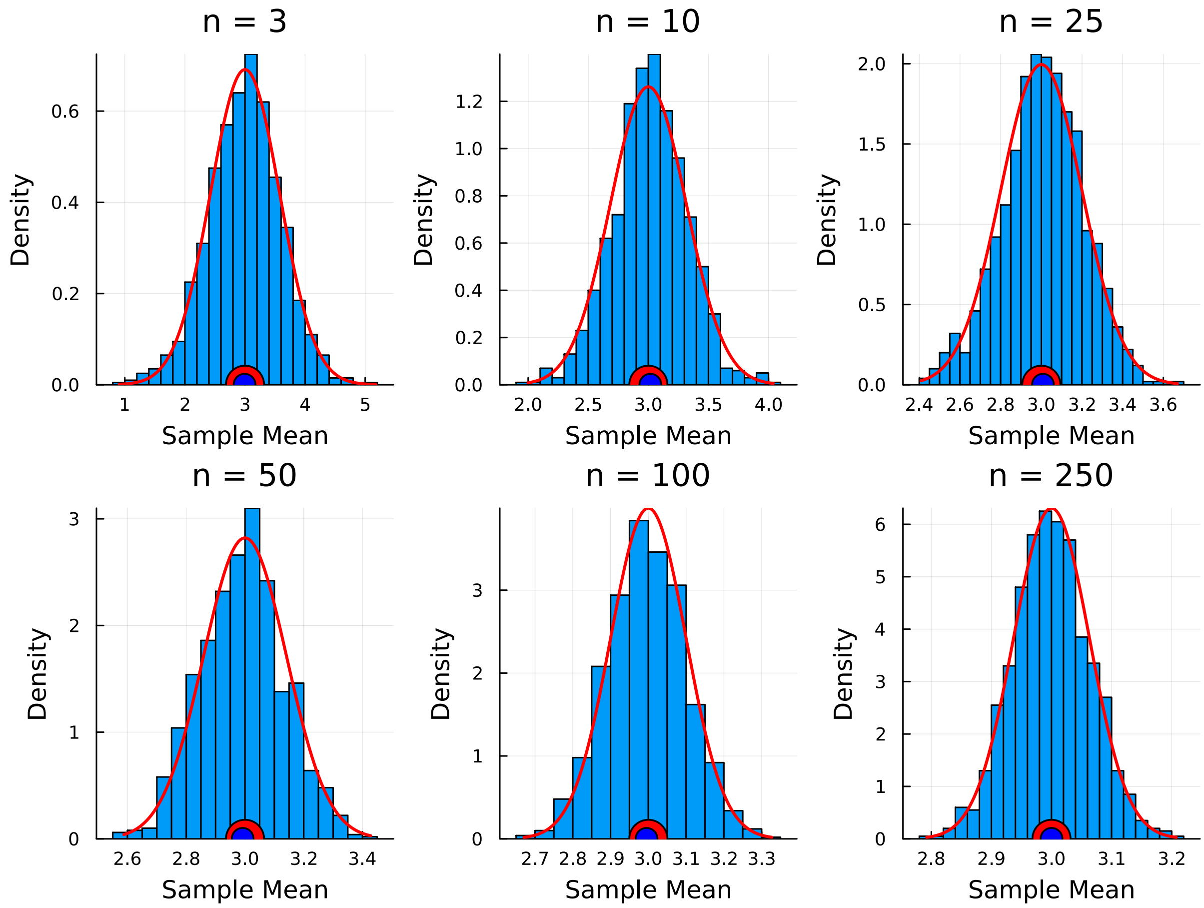

Figure 5: The theoretical PDF of the $\overline{X_n}$ is overlaid on the histograms. The histograms are obtained based on 1000 replications for different sample sizes. For simulation purposes, the population parameters have been fixed at $\mu = 3$ and $\sigma^2 = 1$.

A natural question asked by the students was that how we were actually writing $\mathbb{E}(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$. In fact, the question is much deeper, which states what we really mean by writing

$X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables, each having the same population distribution.

$$X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

From a simulation perspective, if we want to check whether X_1 follows the same population distribution, we need to repeat the sampling $m = 1000$ times (say), and from each sample of size n, we record the first entry, which is a realization of X_1.

In the following, we do this process for $n = 3$. Let us write down the steps in an algorithmic way:

- Fix n (sample size)
- Fix M (number of replications)
- For each $m \in \{1, 2, \ldots, M\}$
  - Simulate $X_1^{(m)}, X_2^{(m)}, X_3^{(m)} \sim \mathcal{N}(\mu, \sigma^2)$
  - Draw histograms of $\{X_j^{(m)}\}_{m=1}^{M}$ for $j \in \{1, 2, 3\}$.
  - Overlay $\mathcal{N}(\mu, \sigma^2)$ on each histogram.
- All three histograms match closely, approximating the population distribution.

In [10]:
```
n = 3
M = 1000
mu = 3
sigma = 1
x1_vals = zeros(M)
x2_vals = zeros(M)
x3_vals = zeros(M)

for i in 1:M
    x = rand(Normal(mu, sigma), n)
    x1_vals[i] = x[1]
    x2_vals[i] = x[2]
    x3_vals[i] = x[3]
end
x_min = minimum([x1_vals; x2_vals; x3_vals]) - 0.5
x_max = maximum([x1_vals; x2_vals; x3_vals]) + 0.5
x_vals = range(x_min, x_max, length=1000)
pdf_normal = pdf.(Normal(mu, sigma), x_vals)


p1 = histogram(x1_vals, normalize= true, xlabel=L"X_1", ylabel="Density",
    title="n = $n", label="")
plot!(p1, x_vals, pdf_normal, color="red", lw=2, label= "")

p2 = histogram(x2_vals, normalize= true, xlabel=L"X_2", ylabel="Density",
    title="n = $n", label="")
plot!(p2, x_vals, pdf_normal, color="red", lw=2, label="")

p3 = histogram(x3_vals, normalize= true, xlabel=L"X_3", ylabel="Density",
    title="n = $n", label="")
plot!(p3, x_vals, pdf_normal, color="red", lw=2, label = "")


plot(p1, p2, p3, layout=(2,2), size=(800, 600))
```

Out[10]:

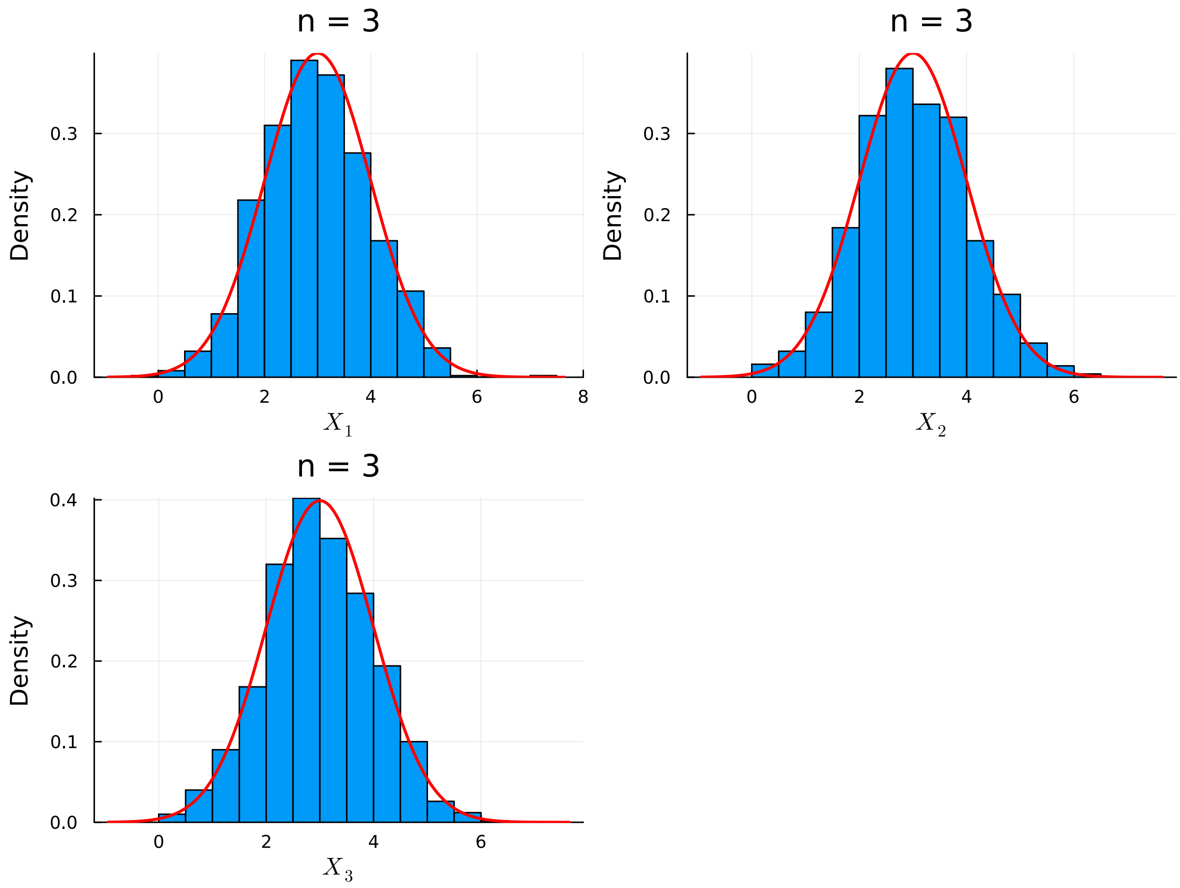

Figure 6: The histogram obtained from the realizations of X_1 is well approximated by the population PDF. This is the same for X_2 and X_3. In addition, you can draw a pairwise scatter plot of these values to check that correlations among them are very close to zero.

Convergence in probability

Suppose that $\{X_n\}$ be a sequence of random variables defined on some probability space $(\mathcal{S}, \mathcal{B}, \mathbb{P})$ and X be another random variable defined on the same probability space. We say that the sequence $\{X_n\}$ converges to X in probability, denoted as $X_n \xrightarrow{P} X$, if for any $\epsilon > 0$ (however small), the following condition holds

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0.$$

Basically for every X_n, we construct a sequence of real numbers (a_n) obtained by

$$a_n = \mathbb{P}(|X_n - X| \geq \epsilon) = \int\int_{\{(x_n,x):|x_n-x|\geq\epsilon\}} f_{X_n,X}(x_n, x)\, dx_n\, dx.$$

If $\lim_{n\to\infty} a_n = 0$, then the sequence of random variables $\{X_n\}$ converges to X in probability.Let us consider a simple example.
Suppose $\{X_n\}$ be a sequence of independent random variables having $\text{Uniform}(0, \theta)$, $\theta \in (0, \infty)$ distribution.Consider the following sequence of random variables

$$Y_n = \max(X_1, X_2, \ldots, X_n)$$

which is also known as the maximum order statistics. We show that $Y_n \to \theta$ in probability. First of all, we compute the sampling distribution of Y_n. We first find the CDF of Y_n and by differentiating the CDF, we can obtain the PDF $f_{Y_n}(y)$.

$$
\begin{aligned}
F_{Y_n}(y) &= P(Y_n \le y) = P(\max(X_1,\ldots,X_n) \le y) \\
&= P(X_1 \le y, X_2 \le y, \ldots, X_n \le y) \\
&= P(X_1 \le y) \cdots P(X_n \le y) \\
&= (P(X_1 \le y))^n = \left(\frac{y}{\theta}\right)^n, \quad 0 < y < \theta.
\end{aligned}
$$

By differentiating the CDF, we obtain the PDF as

$$f_{Y_n}(y) = n\left(\frac{y}{\theta}\right)^{n-1}\frac{1}{\theta}, \quad 0 < y < \theta.$$

and zero, otherwise. Before going to the mathematical proof, let us consider the following simulation exercises. In the following, we obtain the sampling distribution of Y_n based on computer simulation using the following algorithm:

- Fix n.
- Fix θ.
- Fix m, the number of replications.
- Simulate $X_1, X_2, \ldots X_n \sim \text{Uniform}(0,\theta)$.
- Compute $Y_n = \max(X_1, X_2, \ldots, X_n)$.
- Repeat the previous two steps m times to obtain $Y_n^{(1)}, \ldots, Y_n^{(m)}$.
- Draw the histogram of the values $\{Y_n^{(j)}\}_{j=1}^m$).
- Overlay the exact PDF $f_{Y_n}(y)$ on the histogram.

In the above step, at each step $j \in \{1, 2, \ldots, m\}$ we fix $Random.seed(j)$, so that the figures are reproducible. In the following codes, we do the same experiment for minimum order statistics as well which is given by

$$Y_1 = \min(X_1, X_2, \ldots, X_n).$$

In [11]:
```
using Plots, Statistics, StatsPlots, Random
using LaTeXStrings, Distributions
```

In [12]:
```
theta = 1
n = 10
x = rand(Uniform(0, theta),n)
print(maximum(x))
```

```
0.9678401224539817
```

In [13]:
```
m = 1000
y_1 = zeros(m)
y_n = zeros(m)

for i in 1:m
    Random.seed!(i)
    x = rand(Uniform(0, theta),n)
    y_1[i] = minimum(x)
    y_n[i] = maximum(x)
end

p1 = histogram(y_1, normalize = true, xlabel = L"Y_1", bins = 30,
ylabel = L"Density", title = "n = $n", ylims = (0,10), label = "")
plot!(x->n*(1-x/theta)^(n-1)*(1/theta), color = "red", lw = 2,
label = "")
```

```
p2 = histogram(y_n, normalize = true, xlabel = L"Y_n", bins = 30,
ylabel = L"Density", title = "n = $n", ylims = (0,10),label = "")
plot!(x->n*(x/theta)^(n-1)*(1/theta), color = "red", lw = 2,
label = "")

plot(p1, p2 , layout = (1,2), size = (800, 500) )
```

Out[13]:

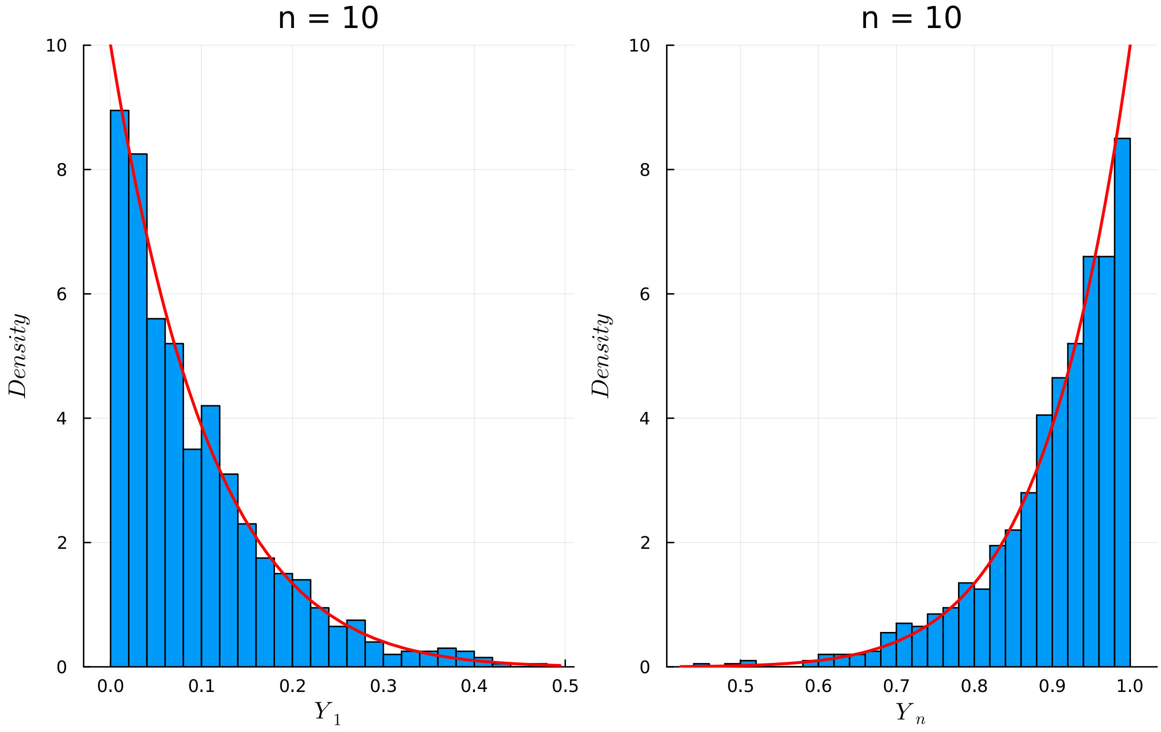

Figure 7: We simulate the sampling distribution of the maximum and minimum order statistics, Y_n and Y_1. The sampling distribution of Y_n is more concentrated towards θ, and Y_1 is more concentrated towards 0. The exact distribution is overlaid on the histograms obtained from simulation studies, and the simulation agrees with the theoretical claim.

The exact sampling distribution of the minimum order statistics Y_1 can be derived in a similar fashion. First, we compute the CDF of Y_1.

$$\begin{aligned}
F_{Y_1}(y) &= P(Y_1 \leq y) = 1 - P(Y_1 > y) \\
&= 1 - P(\min(X_1, X_2, \ldots, X_n) > y) \\
&= 1 - P(X_1 > y, \ldots, X_n > y) \\
&= 1 - \prod_{i=1}^{n} P(X_i > y) \\
&= 1 - (1 - P(X_1 \leq y))^n \\
&= 1 - \left(1 - \frac{y}{\theta}\right)^n, \quad 0 < y < \theta.
\end{aligned}$$

The PDF of Y_1 is given by

$$f_{Y_1}(y) = \begin{cases} n\left(1 - \frac{y}{\theta}\right)^{n-1} \frac{1}{\theta}, & 0 < y < \theta, \\ 0, & \text{Otherwise.} \end{cases}$$

An alternative visualization

In the above explanation, we have fixed the sample size n and simulated the observations from the sampling distribution of Y_1 and Y_n. The process is repeated 1000 times and we check whether the resulting histograms are well approximated by the theoretically derived PDFs. In the following, we vary $n \in \{1, 2, 3, \ldots\}$ and for each n, we simulate a random sample of size n and compute Y_1 and Y_n. Therefore, in this scheme, we do not have any replications. We plot the sequences $\{Y_1^{(1)}, Y_1^{(2)}, \ldots\}$ and $\{Y_n^{(1)}, Y_n^{(2)}, \ldots\}$ values against $n \in \{1, 2, 3, \ldots\}$. This will give an idea, as sample size increases, whether these random quantities converge to some particular values.

In [14]:

```
n_vals = 1:100
y_1 = zeros(length(n_vals))
y_n = zeros(length(n_vals))

for n in n_vals
    x = rand(Uniform(0, theta),n)
    y_1[n] = minimum(x)
    y_n[n] = maximum(x)
end

plot(n_vals, y_n, seriestype=:scatter, color=:red, lw=2,
    xlabel="sample size (n)", ylim=(0,1), ylabel=" ",
    label=L"Y_n", legend_position=(0.5,0.5))
plot!(n_vals, y_1, color=:blue, seriestype=:scatter, lw=2,
    label=L"Y_1")
```

Out[14]:

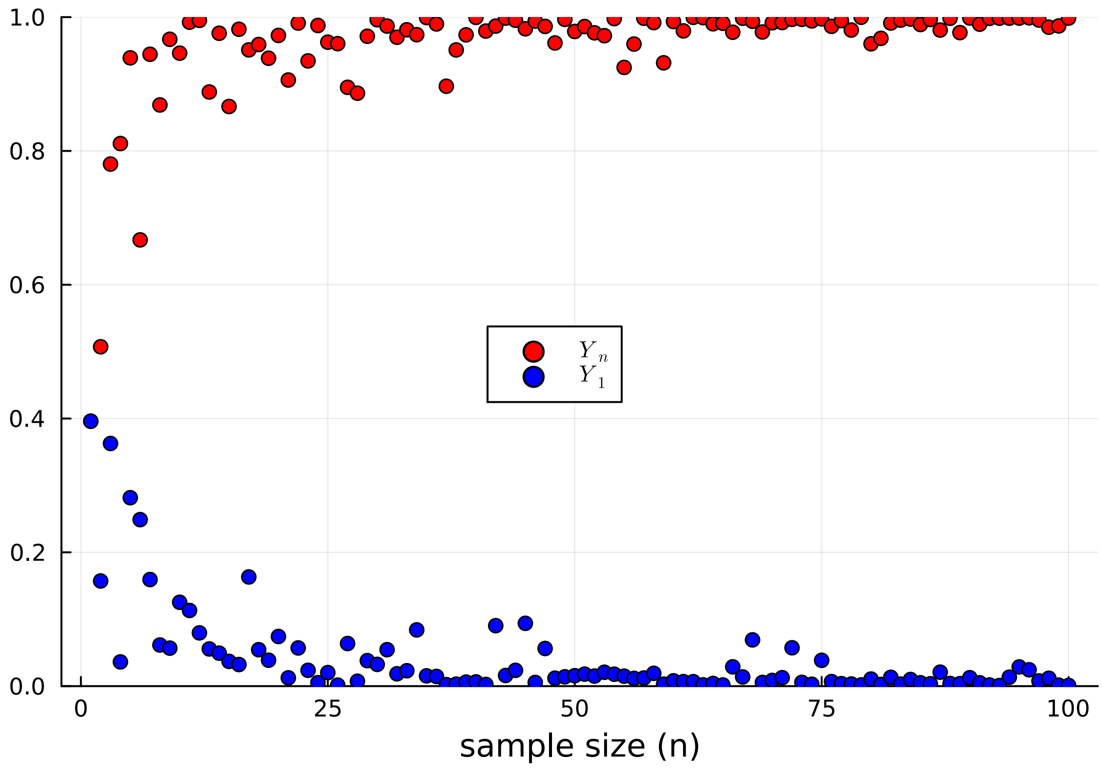

Figure 8: Random sample of size n is simulated from the uniform distribution from Uniform(0, θ). θ = 1 is assumed for simulation. Simulated realization of the minimum and maximum order statistics are shown in the graph. It is observed that $Y_n = \max(X_1, \ldots, X_n) \to 1$ and $Y_1 = \min(X_1, \ldots, X_n) \to 0$ as $n \to \infty$.

Central Limit Theorem

The Central Limit Theorem is one of the most fundamental concepts in Statistics which has profound applications across all domains of Data Science and Machine Learning. It says that the sample mean is approximately normally distributed for large sample size n. However, we shall see it more critically and what are the assumptions needed to understand it in a better way. We shall see some mathematical proof and a lots of simulation to understand this idea better. To understand CLT, first we understand the concept of the convergence in distribution. We elaborate it using a couple of motivating examples along with visualization using Julia.

Convergence in Distributions

Suppose that $X_1, X_2, \ldots, X_n$ be a sequence of random variables following the Uniform $(0, \theta)$ distribution, where $\theta \in (0, \infty) = \Theta$ (say). Suppose that Y_n be the maximum order statistic whose CDF is given by

$$F_n(x) = P(Y_n \leq x) = \begin{cases} 0, & -\infty < x < 0, \\ \left(\frac{x}{\theta}\right)^n, & 0 \leq x < \theta, \\ 1, & \theta \leq x < \infty. \end{cases}$$

We are interested in knowing how the function $F_n(x)$ behaves at each value of x as $n \to \infty$. It is easy to observe that:

- If $x < 0$, then $\lim_{n\to\infty} F_n(x) = 0$.
- If $0 \leq x < \theta$, then $\lim_{n\to\infty} F_n(x) = 0$.
- If $\theta \leq x < \infty$, then $\lim_{n\to\infty} F_n(x) = 1$.

Therefore, the limiting CDF is given by the following function, which represents the **CDF** of a random variable X that takes the value θ with probability 1:

$$F_X(x) = \begin{cases} 1, & \theta \leq x < \infty, \\ 0, & -\infty < x < \theta. \end{cases}$$

In the following, we visualize the convergence graphically. It is important to note that each $F_n(x)$ is a continuous function for $n \in \{1, 2, 3, \ldots\}$, whereas the limiting CDF is not continuous at $x = 0$. We say that the maximum order statistic $Y_n = \max(X_1, X_2, \ldots, X_n)$ converges in distribution to the random variable X.

In [15]:
```
using Plots, Statistics, StatsPlots
using LaTeXStrings, Distributions
```

In [16]:
```
theta = 2
n = 1

function F_n(x, n)
     (x/theta)^n * (0 <= x < theta) + (theta <= x)
end

plot(x -> F_n(x, n), -0.5, 2.5, color = 2, linestyle = :dash,
    lw = 2, ylabel= L"F_n(x)", label="")

n_vals = [3, 5, 10, 25]
for n in n_vals
    plot!(x -> F_n(x, n), -0.5, 2.5, color= n, linestyle = :dash,
        lw =2, label= "n = $n")
```

```
end

scatter!([0, theta], [0, 1], marker=:circle, color= "red",
    markersize = 6, label="")
plot!([-0.5, theta], [0, 0], color= "blue", linewidth=3, label="")
plot!([theta, 2.5], [1, 1], color=:blue, linewidth=3, label="")
```

Out[16]:

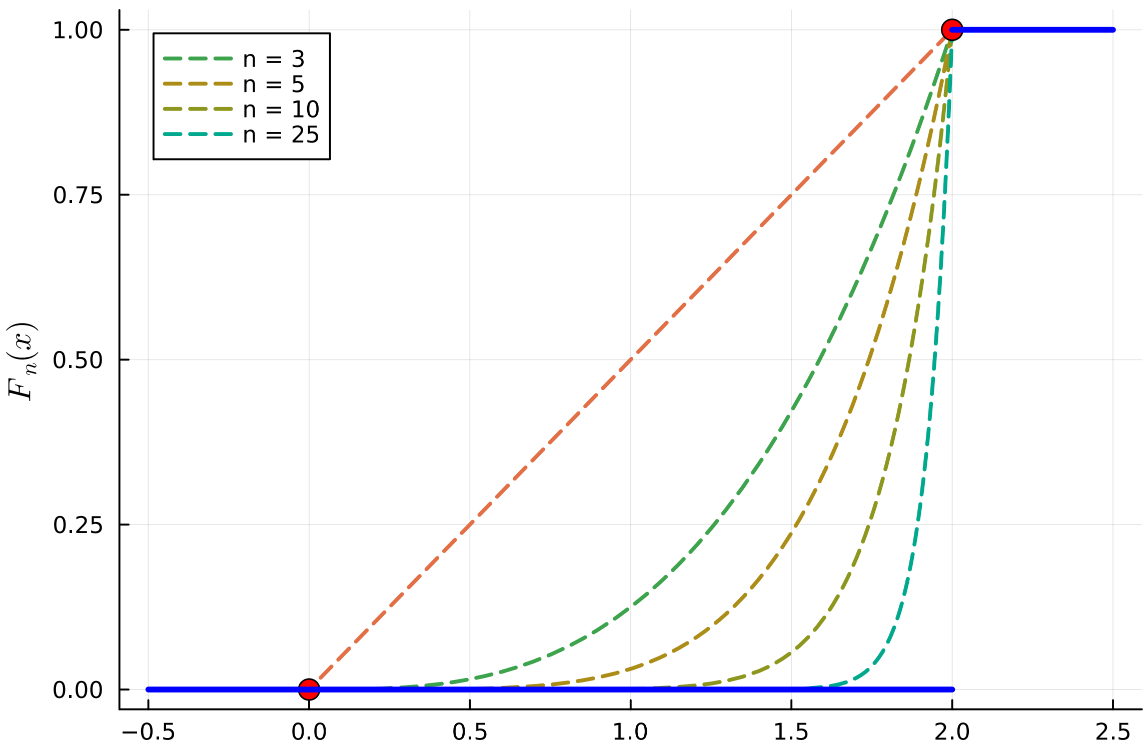

Figure 9: As the sample size increases, the function $F_n(x)$ converges to the function $F_X(x)$ at all points x except 0. However, note that the point $x = 0$ is a point of discontinuity.

In the light of the above example, let us formalize the concept of the convergence in distribution.

> ❗ **Convergence in Distribution**

A sequence of random variables X_n defined on some probability space $(\mathcal{S}, \mathcal{B}, P)$ with CDF $F_n(x)$ is said to converge in distribution to another random variable X having CDF $F_X(x)$, if the following holds

$$\lim_{n\to\infty} F_n(x) = F_X(x)$$

> at all points x where the function $F_X(x)$ is continuous.

An important point to be noted here is that for checking the convergence in distribution, we need to check on the convergence at all the points where the function $F_X(x)$ is continuous. Let us now see another example of convergence in distribution. Suppose that

$$X_n \sim \sqrt{n} \times \text{Uniform}\left(0, \frac{1}{n}\right).$$

Let $U_n \sim \text{Uniform}\left(0, \frac{1}{n}\right)$ and then $X_n = \sqrt{n} \times U_n$.

- U_n shrinks towards 0 as $n \to \infty$.

- It is being amplified again by multiplying $\sqrt{n}$.

Therefore, X_n is basically a product of two components, one is exploding as $n \to \infty$ and another component is shrinking as $n \to \infty$.

- Let us investigate the U_n component as follows through some visualization. The CDF and PDF of U_n is given by

$$f_{U_n}(x) = \begin{cases} n, & 0 < x < \frac{1}{n} \\ 0, & \text{otherwise} \end{cases}$$

$$F_{U_n}(x) = \begin{cases} 0, & -\infty < x < 0 \\ nx, & 0 \leq x < \frac{1}{n} \\ 1, & \frac{1}{n} \leq x < \infty \end{cases}$$

In [17]:
```
n = 1
function f_n(x,n)
     n*(0<x)*(x<1/n)
end

p1 = plot(x -> f_n(x, n), 0, 1, color = 2, linestyle = :dash,
     xlabel = L"x", lw = 2, ylabel= L"f_n(x)", label="")
n_vals = [3, 5, 8, 25]
for n in n_vals
    plot!(x -> f_n(x, n), 0, 1, color= n, linestyle = :dash,
        lw = 2, label= "n = $n")
end

n = 1
function F_n(x,n)
     n*x*(0<x)*(x<1/n) + (1/n<=x)
end

p2 = plot(x -> F_n(x, n), -0.1, 1.1, color = n, linestyle = :dash,
    xlabel = L"x",lw = 2, ylabel= L"F_n(x)", label="")
n_vals = [3, 5, 8, 25]
for n in n_vals
    plot!(x -> F_n(x, n), 0, 1, color= n, linestyle = :dash,
        lw = 2, label= "n = $n")
end

scatter!([0], [1], marker=:circle, color= "red",
    markersize = 5, label="")
plot!([-0.1, 0, 0, 1.5], [0, 0, 1, 1], color= "blue",
    linewidth= 3, label="")

plot(p1, p2, layout = (1,2), size = (800,400))
```

Out[17]:

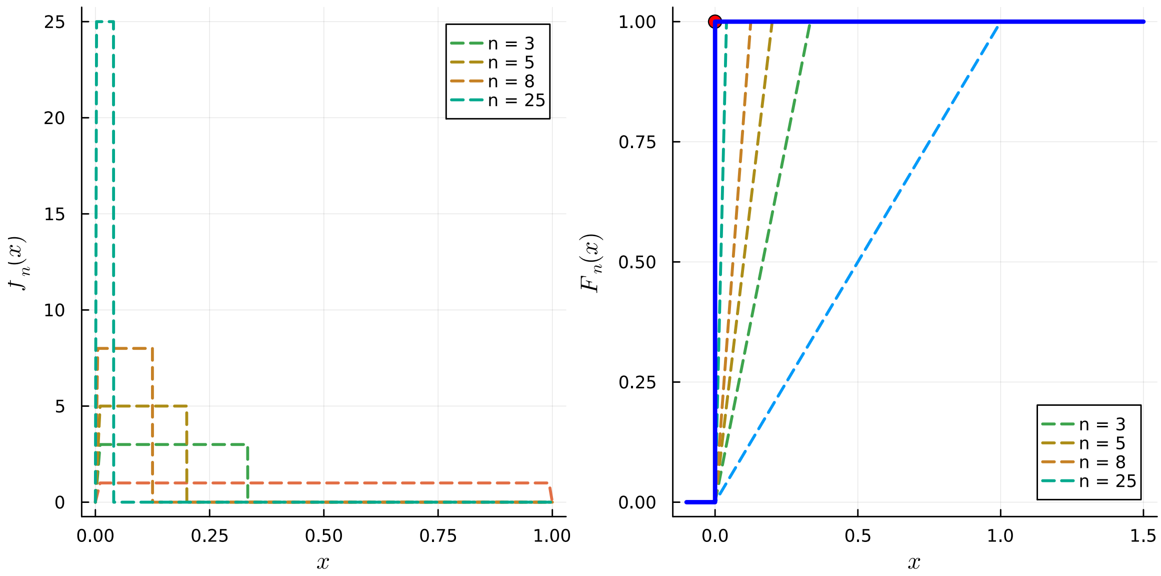

Figure 10: As $n \to \infty$, $f_n(x)$ becomes highly concentrated at 0 and $F_n(x)$ approaches to the CDF of a random variable X which takes 0 with probability 1.

Therefore, U_n converges in distribution to 0 and we are now ready to check that if we multiply $\sqrt{n}$ with U_n, will it still go to zero? In the following let us consider the simulation of U_n for different values of n and also compute X_n and see how these simulated numbers behave as $n \to \infty$.

In [18]:

```
n = 1000
U_n = zeros(n)

for i in 1:n
    U_n[i] = rand(Uniform(0, 1/i))
end

plot(1:n, U_n, color = "blue", xlabel = "n", lw = 2,
    label = L"U_n")

X_n = zeros(n)
for i in 1:n
    X_n[i] = sqrt(i).*U_n[i]
end

plot!(1:n, X_n, color = "red", lw = 2, label = L"X_n")
```

Out[18]:

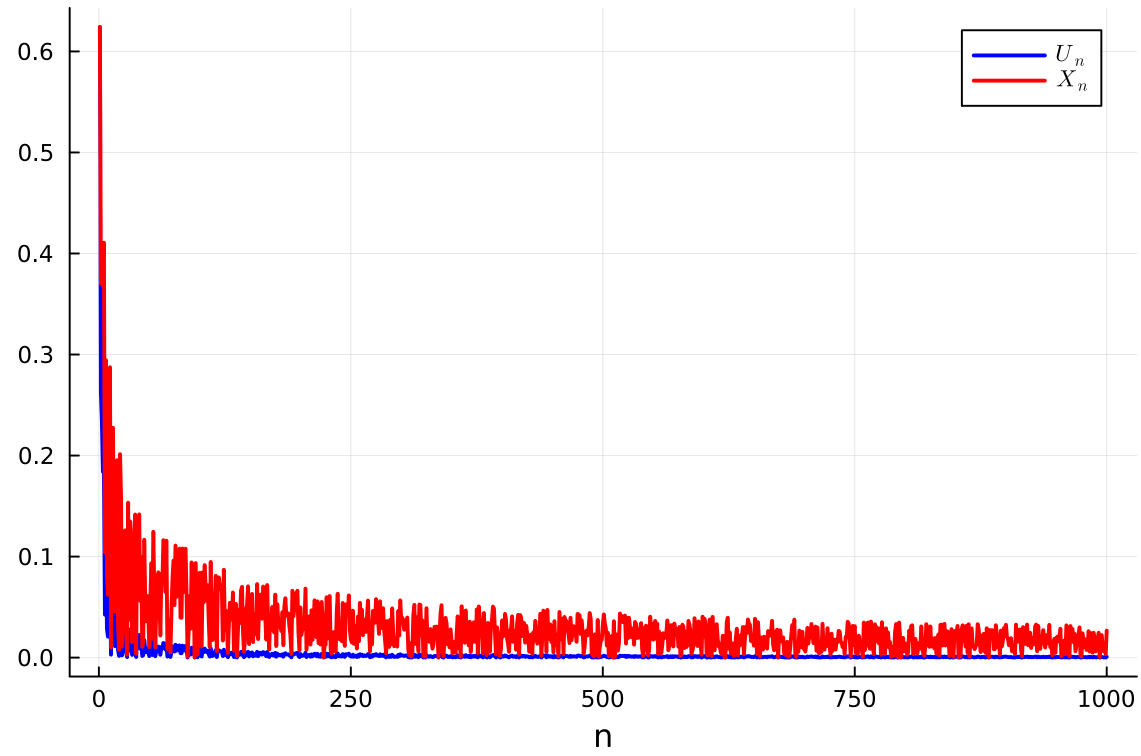

Figure 11: Simulations suggest that both U_n and X_n converges to 0, but convergence of X_n to 0 is slower. Convergence of U_n and X_n are shown in blue and red colour only.

We can expand the scope of this problem by expanding the definition of X_n as follows:

$$X_n = n^\delta U_n, \quad \delta > 0.$$

In the following, we choose different values of δ and see how the simulations behave. The following simulations suggest that

- $0 < \delta < 1$, $X_n \to 0$ in probability and distribution as well.
- $\delta = 1$, $X_n \to \text{Uniform}(0, 1)$ as $n \to \infty$.
- $\delta > 1$, X_n does not converge as $n \to \infty$.

In [19]:

```
delta_vals = [0.1, 0.3, 0.5, 0.8, 1, 1.2]

plt = plot(layout=(2, 3), size=(800, 400))

for (idx, delta) in enumerate(delta_vals)
    n = 1000
    U_n = zeros(n)
    for i in 1:n
          U_n[i] = rand(Uniform(0, 1/i))
    end
    X_n = zeros(n)
    for i in 1:n
         X_n[i] = (i).^delta.*U_n[i]
    end
    plot!(1:n, X_n, color = "red", xlabel = "n", lw = 2,
    label = L"X_n", title = L"\delta = %$delta", subplot = idx)
    plot!(1:n,U_n, color = "blue", label = L"U_n", subplot = idx)
end

display(plt)
```

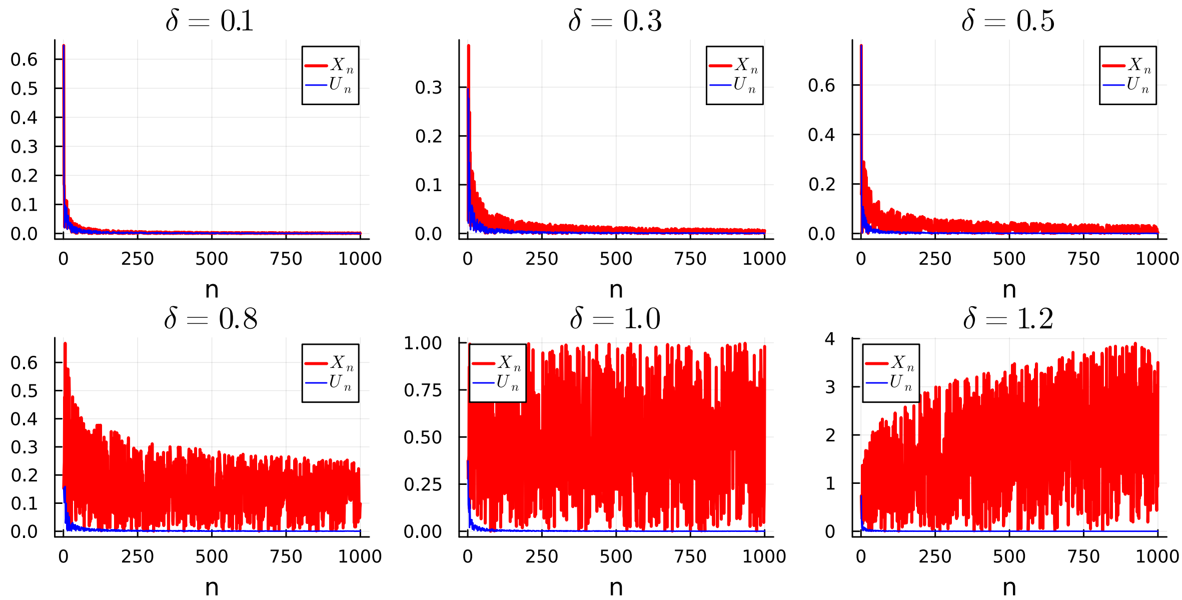

Figure 12: The convergence of X_n is shown for different choices of δ values. As δ increases (but less than 1), the rate of convergence to 0 is slower. At $\delta = 1$, it converges to Uniform(0,1) and for $\delta > 1$, the sequence diverges.

We have learnt two different convergence concepts: Convergence in probability and the convergence in distribution. In notation they are denoted as

$$X_n \xrightarrow{P} X, \quad X_n \xrightarrow{d} X.$$

The following results hold:

- If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$. The converse of the statement is not true in general. However, if $X_n \xrightarrow{P} X$ and X is a degenerate random variable (that is a constant), then $X_n \xrightarrow{P} X$ as well.
- If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n + Y_n \xrightarrow{P} X + Y$.
- If $X_n \xrightarrow{P} X$ and a is a real number, then $aX_n \xrightarrow{P} aX$.
- If (a_n) be a sequence of real numbers with $a_n \to a$ as $n \to \infty$ and $X_n \xrightarrow{P} X$, then $a_n X_n \xrightarrow{P} aX$.

Insight into the Central Limit Theorem

Suppose that we have a random sample of size n from the population distribution whose PDF or PMF is given by $f(x)$. In addition, assume that the population has finite variance $\sigma^2 < \infty$. Let $\mu = \mathbb{E}(X)$ be the population mean. We denote the sample mean using the symbol

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

Due to the Weak Law of Large Numbers (WLLN), we have observed that $\overline{X}_n \to \mu$ in probability as $n \to \infty$. It is to be noted that for every $n \geq 1$, $\overline{X}_n$ is a random variable and depending on the problem, its exact probability distribution may be computed. For example, if $X_1, X_2, \dots, X_n$ be a random sample of size n from the normal distribution with mean μ and variance σ^2, then

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{for all } n \geq 1.$$

One can show that the Moment Generating Function (MGF) of $\overline{X}_n$ is given by

$$M_{\overline{X}_n}(t) = e^{\mu t + \frac{1}{2}\frac{\sigma^2}{n}t^2}, \quad -\infty < t < \infty.$$

which establishes the result by the uniqueness of the MGF.

Now, can we generalize this result for any population distribution with finite variance? In particular, can for large sample size n, the sampling distribution of $\overline{X}_n$ be approximated by some normal distribution? If so, what will be the mean and variance of the approximating normal distribution?
Before going to provide an exact answer to the above questions, let us perform some computer simulations and see how the sampling distribution of the sample mean behaves for large sample size n and for different population distributions.

Experiment with exponential

In the following code, we repeatedly simulate a sample of size n from the exponential distribution with rate λ, which is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 < x < \infty \\ 0, & \text{otherwise} \end{cases}.$$

The distribution is positively skewed and does not have any apparent connection with the normal distribution. We implement the following algorithm:

- Fix sample size n
- Fix the population parameter λ
- Fix the number of replications m
- Simulate $X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exponential}(\lambda)$
- Compute the sample mean $\overline{X}_n$
- Repeat the previous two steps m times to obtain m realizations $\left\{\overline{X}_n^{(j)}\right\}_{j=1}^{m}$ from the distribution of $\overline{X}_n$.
- Plot the histogram of $\left\{\overline{X}_n^{(j)}\right\}_{j=1}^{m}$ values.
- Overlay $\mathcal{N}\left(\frac{1}{\lambda}, \frac{1}{\lambda^2 n}\right)$ to the histogram.
- Repeat the above exercises for different (in increasing order) of sample size n.

In [20]:
```
lambda = 2
n_vals = [3,5,10,25,50,100]
rep = 1000
plt = plot(layout=(2, 3), size=(800, 600))
```

```
for (idx, n) in enumerate(n_vals)
    sample_mean = zeros(rep)
    for i in 1:rep
        x = rand(Exponential(1/lambda), n)
        sample_mean[i] = mean(x)
    end
    histogram!(sample_mean, normalize = true, bins = 30,
    xlabel= L"\bar{X}", ylabel = "Density", title = "n = $n",
    label = "", subplot = idx)

    x_vals = range(minimum(sample_mean), maximum(sample_mean),
        length = 1000)
    pdf_normal = pdf.(Normal(1/lambda, sqrt(1/(lambda^2 * n))),
        x_vals)
    pdf_gamma = pdf.(Gamma(n, 1/(lambda * n)), x_vals)
    plot!(x_vals, pdf_normal, color = "red", lw = 2, label = "",
    subplot = idx)
    plot!(x_vals, pdf_gamma, color = "blue", lw = 2, label = "",
    subplot = idx)
end

display(plt)
```

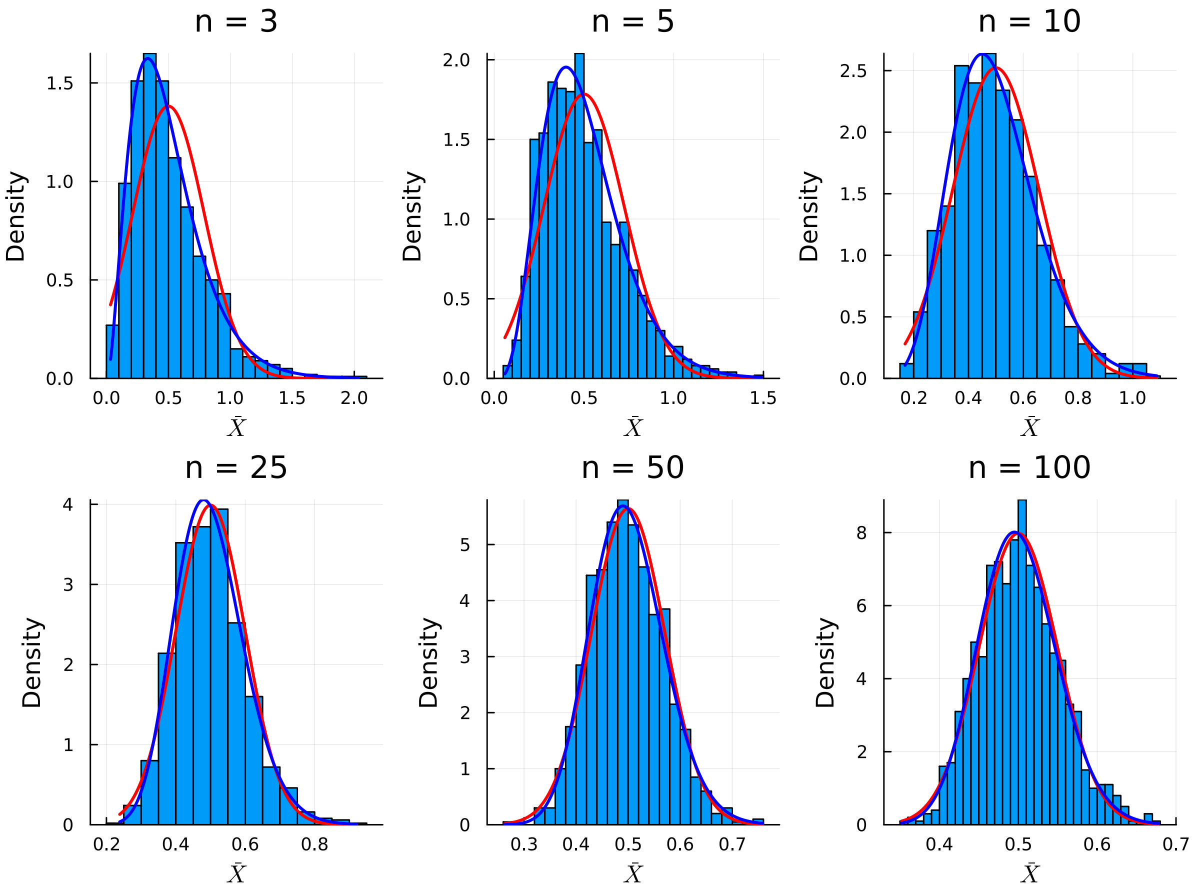

Figure 13: As the sample size increases, the sampling distribution of the sample mean can be well approximated by the normal distribution with mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2 n}$. The histograms are obtained based on 1000 replications from the exponential distribution with rate $\lambda = 2$.

In this simulation study, we are lucky that the exact sampling distribution of the sample mean $\overline{X}_n$ can be derived using the MGF. The sampling distribution is given by

$$f_{\overline{X}_n}(x) = \begin{cases} \dfrac{(n\lambda)^n \, e^{-n\lambda x} \, x^{n-1}}{\Gamma(n)}, & 0 < x < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Experiment with Poisson

We repeat the above simulation exercise for the Poisson distribution with rate λ. In the resulting histograms, we overlay the normal distribution with mean λ and variance $\frac{\lambda}{n}$.

In [21]:

```
lambda = 2
n_vals = [3,5,10,25,50,100]
rep = 1000
plt = plot(layout=(2, 3), size=(800, 600))

for (idx, n) in enumerate(n_vals)
    sample_mean = zeros(rep)
    for i in 1:rep
        x = rand(Poisson(lambda), n)
        sample_mean[i] = mean(x)
    end
    histogram!(sample_mean, normalize = true, bins = 30,
    xlabel= L"\bar{X}", ylabel = "Density", title = "n = $n",
    label = "", subplot = idx)
    x_vals = range(minimum(sample_mean), maximum(sample_mean),
        length = 1000)
    pdf_normal = pdf.(Normal(lambda,sqrt(lambda/n)),
        x_vals)
    plot!(x_vals, pdf_normal, color = "red", lw = 2, label = "",
    subplot = idx)
end

display(plt)
```

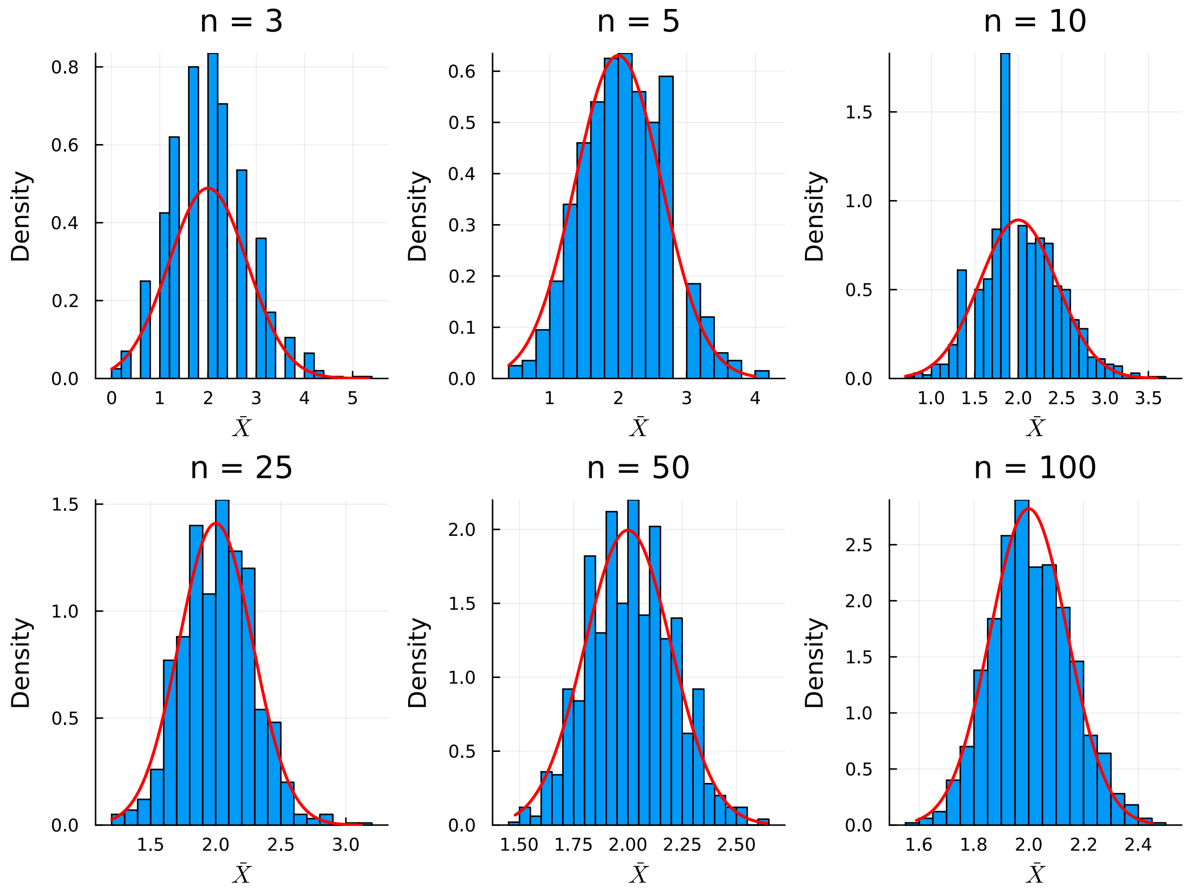

Figure 14: As the sample size increases, the sampling distribution of the sample mean can be well approximated by the normal distribution with mean λ and variance $\frac{\lambda}{n}$. The histograms are obtained based on 1000 replications from the Poisson distribution with rate $\lambda = 2$.

A natural question arises whether the Central Limit Theorem holds for all probability distributions. The answer is negative. Let us consider a simulation experiment using the standard Cauchy distribution, which is given by

$$f(x) = \frac{1}{\pi\left(1 + x^2\right)}, \quad -\infty < x < \infty.$$

Experiment with the Cauchy distribution

In [22]:

```
x = range(-5, 5, length=1000)
y_cauchy = pdf.(Cauchy(0,1), x)
y_norm   = pdf.(Normal(0,1), x)

p1 = plot(x, y_cauchy, color="red", linewidth=2, label="")
plot!(p1, x, y_norm, color="blue", linewidth=2, label="")
n_vals = 1:1000
sample_mean = similar(float.(n_vals))

for (i, n) in enumerate(n_vals)
    sample_mean[i] = mean(rand(Cauchy(0,1), n))
end

p2 = plot(n_vals, sample_mean, color="red", linewidth=2,
          xlabel="Sample size (n)", ylabel="Sample mean",
          title="", label = "")
```

```julia
n = 500
M = 1000
sample_mean = zeros(M)

for i in 1:M
    sample_mean[i] = mean(rand(Cauchy(0,1), n))
end

p3 = histogram(sample_mean, normalize= true, bins = :auto,
    xlims = extrema(sample_mean), title="n = $n", label="")
plot(p1, p2, p3, layout=(2,2), size=(900, 500))
```

Out[22]:

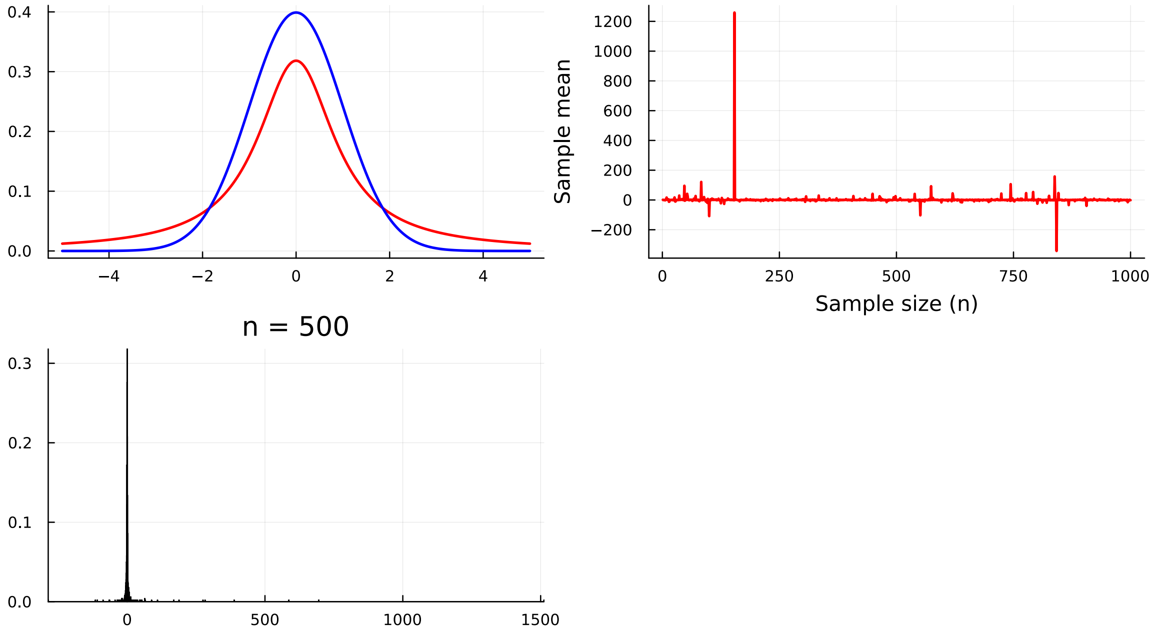

Figure 15: As sample size increases, the sample mean does not converge to μ, the location parameter. Also, for large enough n, the sampling distribution of the sample mean does not converge to a normal distribution. Therefore, for the Cauchy distribution, the WLLN and CLT both do not hold.

> ❗ *WLLN and CLT: Points to Remember*
>
> Based on our discussion, we conclude the following:
>
> - If the population has finite mean, then **WLLN** holds.
> - If the population has finite variance, then **CLT** holds.
> - If the population has finite mean, but the second-order moment does not exist, then **WLLN** will hold but **CLT** will not.
> - If the population has finite variance, then both **WLLN** and **CLT** hold true.

A further generalization of the CLT leads us to investigate the sampling distribution of the function of the sample mean, $g(\overline{X}_n)$, for some function g. For a reasonable choice of the function, can we have some large-sample approximation of the sampling distribution of $g(\overline{X}_n)$?

The Delta Method

Suppose that we have a random sample of size n from the Poisson distribution with rate parameter λ. We are interested in estimating the probability $\mathbb{P}(Y \geq 1)$. In other words, suppose that the number of telephone calls received by a customer in a day is assumed to follow the Poisson distribution with mean λ. We are interested in approximating the probability that the customer will receive at least one call. The MLE of λ is $\overline{X}_n$, the sample mean.

1. **Weak Law of Large Numbers (WLLN)**

$$\overline{X}_n \xrightarrow{p} \lambda,$$

meaning $\overline{X}_n$ converges to the population mean λ in probability as $n \to \infty$. Hence, $\overline{X}_n$ is a consistent estimator of λ.

2. **Central Limit Theorem (CLT)**

For large n,

$$\overline{X}_n \sim \mathcal{N}\!\left(\lambda, \tfrac{\lambda}{n}\right),$$

so the sampling distribution of the sample mean can be well approximated by a normal distribution when the underlying population distribution has finite variance.
We aim to estimate the function

$$\psi(\lambda) \;=\; \mathbb{P}(X \geq 1) \;=\; 1 - e^{-\lambda}.$$

A natural choice is to approximate $\psi(\lambda)$ by substituting the MLE $\overline{X}_n$ into the function:

$$\psi(\overline{X}_n) \;=\; 1 - e^{-\overline{X}_n}.$$

Now our task is to study the behavior of $\psi(\overline{X}_n)$ — including unbiasedness, consistency, and the asymptotic distribution for large n. Even if we are not able to obtain the exact sampling distribution for large n, it is important to check whether we can have some large-sample approximations by known distributions.
Using a first-order Taylor expansion about λ (and neglecting higher-order terms), we get:

$$\psi(\overline{X}_n) \approx \psi(\lambda) + (\overline{X}_n - \lambda)\,\psi'(\lambda)$$

Taking expectations on both sides, we see that

$$E\big[\psi(\overline{X}_n)\big] \approx \psi(\lambda).$$

This means that $\psi(\overline{X}_n)$ is an approximately unbiased estimator of $\psi(\lambda)$ at least for large n.

From the above computation, after some rearranging

$$E\big[\psi(\overline{X}_n) - \psi(\lambda)\big]^2 \approx (\psi'(\lambda))^2\, E\big[(\overline{X}_n - \lambda)^2\big] = (\psi'(\lambda))^2\, \mathrm{Var}(\overline{X}_n) = \frac{\lambda e^{-2\lambda}}{n}.$$

Therefore, the approximate variance of $\psi(\overline{X}_n)$ is

$$\frac{\lambda e^{-2\lambda}}{n}.$$

In the following simulation, we visualize the sampling distribution of $\psi(\overline{X}_n)$.

In [23]:
```
using Plots, StatsBase, StatsModels, Statistics
using Distributions, LaTeXStrings
```

In [24]:
```
plt = plot(layout=(2, 3), size=(800, 500))
n_vals = [4, 10, 20, 50, 100, 500]
rep = 1000
lambda = 2
psi = 1-(1+lambda)*exp(-lambda)

for (idx, n) in enumerate(n_vals)
    psi_vals = zeros(rep)
    for i in 1:rep
        x = rand(Poisson(lambda),n)
        psi_vals[i] = 1-(1+mean(x))*exp(-mean(x))
    end
    histogram!(psi_vals, normalize = true, xlabel = L"\psi",
    ylabel = "Density", title = "n = $n", label = "",
    subplot = idx)
    scatter!([psi],[0], color = "red", markersize = 6, label = ""
    , subplot = idx)
    x_vals = range(minimum(psi_vals), maximum(psi_vals), length = 1000)
    pdf_normal = pdf.(Normal(psi,
             sqrt(lambda^3*exp(-2*lambda)/n)), x_vals)
    plot!(x_vals, pdf_normal, color = "red", lw = 2, label = "",
    subplot = idx)
end

display(plt)
```

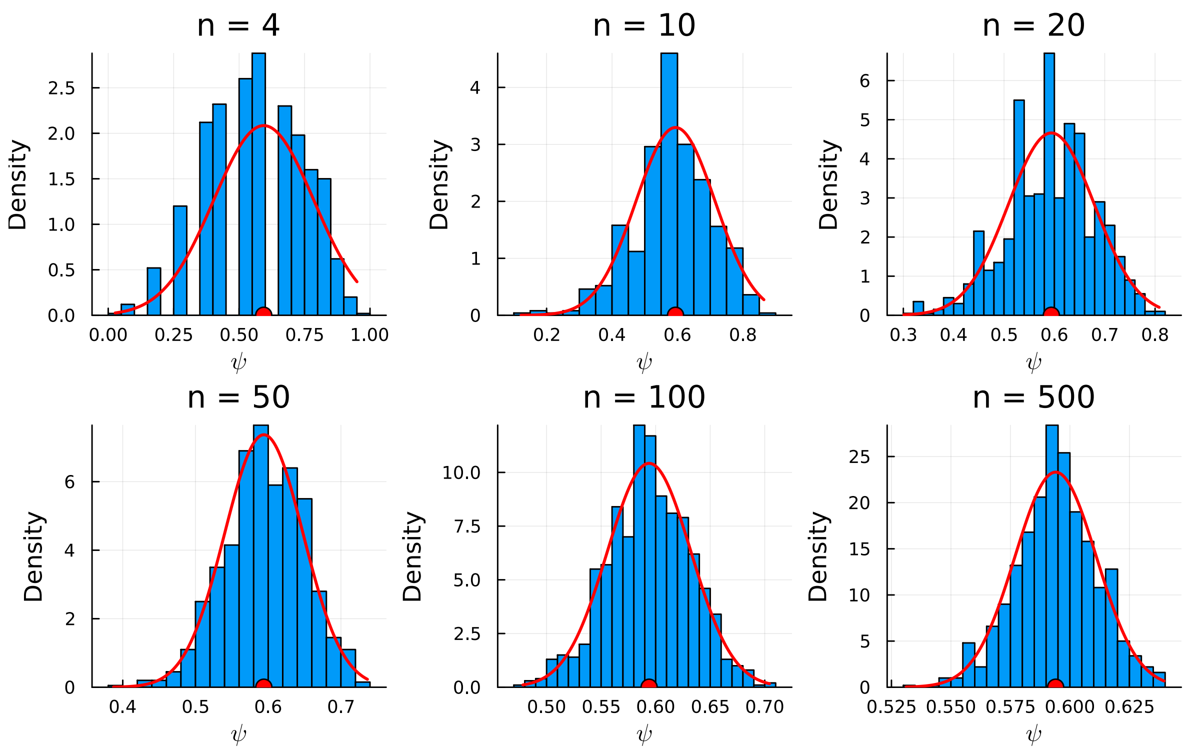

Figure 16: The sampling distribution of the function of the sample mean $\overline{X}_n$ is visualized for different sample sizes by histograms based on 1000 replications.