STATS 200: Introduction to Statistical Inference

Lecture 4: Asymptotics and simulation

Recap

We've discussed a few examples of how to determine the distribution of a **statistic** computed from data, assuming a certain probability model for the data.

For example, last lecture we showed the following results: If $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0, 1)$, then

$$\bar{X} \sim \mathcal{N}\left(0, \frac{1}{n}\right),$$

$$X_1^2 + \ldots + X_n^2 \sim \chi_n^2.$$

Reality check

For many (seemingly simple) statistics, it's difficult to describe its PMF or PDF exactly. For example:

1. Suppose $X_1, \ldots, X_{100} \overset{IID}{\sim} \text{Uniform}(-1, 1)$. What is the distribution of $\bar{X}$?
2. Suppose $(X_1, \ldots, X_6) \sim \text{Multinomial}\left(500, \left(\frac{1}{6}, \ldots, \frac{1}{6}\right)\right)$. What is the distribution of

$$T = \left(\frac{X_1}{500} - \frac{1}{6}\right)^2 + \ldots + \left(\frac{X_6}{500} - \frac{1}{6}\right)^2?$$

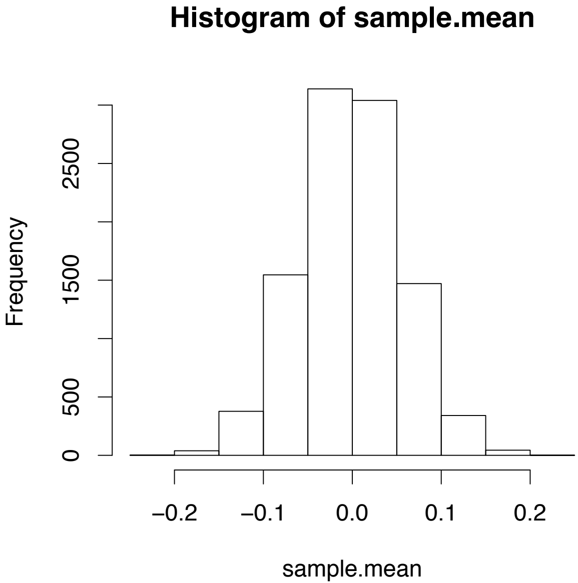
For questions that we don't know how to answer exactly, we'll try to answer them approximately.

Sample mean of IID uniform

If we fully specify the distribution of data, then we can always **simulate** the distribution of any statistic:

```
nreps = 10000
sample.mean = numeric(nreps)
n = 100
for (i in 1:nreps) {
    X = runif(n, min=-1, max=1)
    sample.mean[i] = mean(X)
}
hist(sample.mean)
```

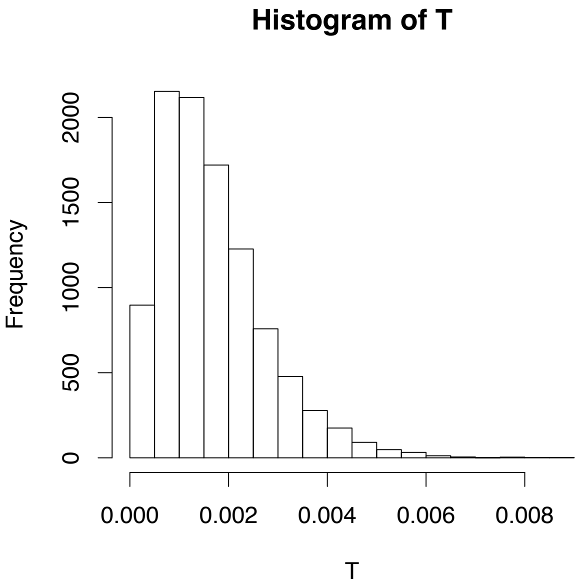
Sample mean of IID uniform



Is your friend cheating you in dice?

```
nreps = 10000
T = numeric(nreps)
n = 500
p = c(1/6,1/6,1/6,1/6,1/6,1/6)
for (i in 1:nreps) {
    X = rmultinom(1,n,p)
    T[i] = sum((X/n-p)^2)
}
hist(T)
```

Is your friend cheating you in dice?



Asymptotic analysis

Oftentimes, a very good approximate answer emerges when n is large (in other words, you have many samples). We call results that rely on this type of approximation **asymptotic**.

If we can just simulate, why do asymptotic analysis?

1. Better understanding of the behavior. (Understanding the assumptions: What if X_i are not uniform? What if I don't really know the distribution of X_i? Understanding the scaling: What if $n = 1000$ instead of 100? What if $n = 1{,}000{,}000$?)
2. Faster to get an answer.

(Weak) Law of Large Numbers

Theorem (LLN)

Suppose $X_1, \ldots, X_n$ *are IID, with* $\mathbb{E}[X_1] = \mu$ *and* $\mathrm{Var}[X_1] < \infty$. *Let* $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$. *Then, for any fixed* $\varepsilon > 0$, *as* $n \to \infty$,

$$\mathbb{P}[|\bar{X}_n - \mu| > \varepsilon] \to 0.$$

(Weak) Law of Large Numbers

Theorem (LLN)

Suppose $X_1, \ldots, X_n$ *are IID, with* $\mathbb{E}[X_1] = \mu$ *and* $\mathrm{Var}[X_1] < \infty$*. Let* $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$*. Then, for any fixed* $\varepsilon > 0$*, as* $n \to \infty$*,*

$$\mathbb{P}[|\bar{X}_n - \mu| > \varepsilon] \to 0.$$

A sequence of random variables $\{T_n\}_{n=1}^{\infty}$ **converges in probability** to a constant $c \in \mathbb{R}$ if, for any fixed $\varepsilon > 0$, as $n \to \infty$,

$$\mathbb{P}[|T_n - c| > \varepsilon] \to 0.$$

So the LLN says $\bar{X}_n \to \mu$ in probability.

Central Limit Theorem

Theorem (CLT)

Suppose $X_1, \ldots, X_n$ *are IID, with* $\mathbb{E}[X_1] = \mu$ *and* $\mathrm{Var}[X_1] = \sigma^2 < \infty$*. Let* $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$*. Then, for any fixed* $x \in \mathbb{R}$*, as* $n \to \infty$*,*

$$\mathbb{P}\left[\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \leq x\right] \to \Phi(x),$$

where Φ *is the CDF of the* $\mathcal{N}(0, 1)$ *distribution.*

Central Limit Theorem

Theorem (CLT)

Suppose $X_1, \ldots, X_n$ *are IID, with* $\mathbb{E}[X_1] = \mu$ *and* $\mathrm{Var}[X_1] = \sigma^2 < \infty$. *Let* $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$. *Then, for any fixed* $x \in \mathbb{R}$, *as* $n \to \infty$,

$$\mathbb{P}\left[\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \leq x\right] \to \Phi(x),$$

where Φ *is the CDF of the* $\mathcal{N}(0,1)$ *distribution.*

$\{T_n\}_{n=1}^{\infty}$ **converges in distribution** to a probability distribution with CDF F if, for every $x \in \mathbb{R}$ where F is continuous, as $n \to \infty$,

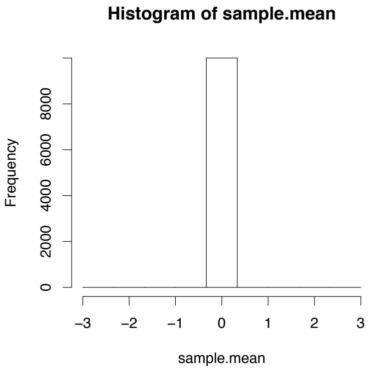
$$\mathbb{P}[T_n \leq x] \to F(x).$$

We sometimes write $T_n \to Z$ in distribution, where Z is a random variable having this distribution F. So the CLT says $\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \to Z$ in distribution where $Z \sim \mathcal{N}(0,1)$.

The Difference is in Scaling

How can the same statistic $\bar{X}_n$ converge both in probability and in distribution? The difference is in scaling:

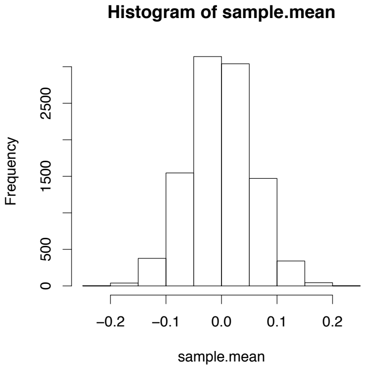
$X_1, \ldots, X_{100} \sim \text{Uniform}(-1, 1)$. $\bar{X}_{100}$ across 10000 simulations:



This illustrates the LLN, that is, $\bar{X}_n \to 0$ in probability.

The Difference is in Scaling

Here's the exact same histogram, on a different scale:



This illustrates the CLT, that is, $\sqrt{3n}\bar{X}_n \to \mathcal{N}(0,1)$ in distribution. (Here $\text{Var}[X_1] = \frac{1}{3}$.)

Sample mean of IID uniform

By the CLT, the distribution of $\bar{X}_n$ is approximately $\mathcal{N}\left(0, \frac{1}{3n}\right)$.

How good is this approximation? Here's a comparison of CDF values, for sample size $n = 10$:*

Normal	Exact
0.01	0.009
0.25	0.253
0.50	0.500
0.75	0.747
0.99	0.991

It's already very close! In general, accuracy depends on

- Sample size n,
- Skewness of the distribution of X_i, and
- Heaviness of tails of the distribution of X_i

*Using `www.math.uah.edu/stat/apps/SpecialCalculator.html`

Multivariate generalizations

Consider

$$\mathbf{X} = (X_1, \ldots, X_k) \in \mathbb{R}^k$$

(with some k-dimensional joint distribution), and let

$$\mu_i = \mathbb{E}[X_i],\ \Sigma_{ii} = \mathrm{Var}[X_i],\ \Sigma_{ij} = \mathrm{Cov}[X_i, X_j].$$

Let $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)} \in \mathbb{R}^k$ be IID, each with the same joint distribution as $\mathbf{X}$. Let $\bar{\mathbf{X}}_n = \frac{1}{n}(\mathbf{X}^{(1)} + \ldots + \mathbf{X}^{(n)}) \in \mathbb{R}^k$.

For example: We measure the height and weight of n randomly chosen people. $\mathbf{X}^{(l)} \in \mathbb{R}^2$ is the height and weight of person l. Height is not independent of weight for the same person, but let's assume they are IID across different people. $\bar{\mathbf{X}}_n \in \mathbb{R}^2$ is the average height and average weight of the n people.

Multivariate generalizations

Theorem (LLN)

As $n \to \infty$, $\bar{\mathbf{X}}_n$ converges in probability to μ.

Theorem (CLT)

As $n \to \infty$, $\sqrt{n}(\bar{\mathbf{X}}_n - \mu)$ converges in distribution to the multivariate normal distribution $\mathcal{N}(0, \Sigma)$.

(We say a sequence $\{\mathbf{T}_n\}_{n=1}^{\infty}$ of random vectors in $\mathbb{R}^k$ converges in probability to $\mu \in \mathbb{R}^k$ if $\mathbb{P}[\|\mathbf{T}_n - \mu\| > \varepsilon] \to 0$ for any $\varepsilon > 0$, where $\|\cdot\|$ is the vector length. We say $\{\mathbf{T}_n\}_{n=1}^{\infty}$ converges in distribution to $\mathbf{Z}$ if, for any set $A \subseteq \mathbb{R}^k$ such that $\mathbf{Z}$ belongs to the boundary of A with probability 0, $\mathbb{P}[\mathbf{T}_n \in A] \to \mathbb{P}[\mathbf{Z} \in A]$.)

Approximating the multinomial distribution for large n

Suppose $(Y_1, \ldots, Y_6) \sim \text{Multinomial}\left(n, \left(\frac{1}{6}, \ldots, \frac{1}{6}\right)\right)$. Y represents the number of times we obtain 1 through 6 when rolling a 6-sided die n times.

For each $l = 1, \ldots, n$, let $\mathbf{X}^{(l)} = (1, 0, 0, 0, 0, 0)$ if we got 1 on the l^{th} roll, $(0, 1, 0, 0, 0, 0)$ if we got 2 on the l^{th} roll, etc. Then $(Y_1, \ldots, Y_6) = \mathbf{X}^{(1)} + \ldots + \mathbf{X}^{(n)}$.

Let's apply the (multivariate) LLN and CLT!

Approximating the multinomial distribution for large n

Let's write $\mathbf{X}^{(1)} = (X_1, \ldots, X_6)$, so $X_1, \ldots, X_6$ are random variables where exactly one them equals 1 (and the rest equal 0). Then:

$$\mathbb{E}[X_i] = \mathbb{P}[X_i = 1] = \frac{1}{6},$$

Approximating the multinomial distribution for large n

Let's write $\mathbf{X}^{(1)} = (X_1, \ldots, X_6)$, so $X_1, \ldots, X_6$ are random variables where exactly one them equals 1 (and the rest equal 0). Then:

$$\mathbb{E}[X_i] = \mathbb{P}[X_i = 1] = \frac{1}{6},$$

$$\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = \frac{1}{6} - \left(\frac{1}{6}\right)^2 = \frac{5}{36},$$

Approximating the multinomial distribution for large n

Let's write $\mathbf{X}^{(1)} = (X_1, \ldots, X_6)$, so $X_1, \ldots, X_6$ are random variables where exactly one them equals 1 (and the rest equal 0). Then:

$$\mathbb{E}[X_i] = \mathbb{P}[X_i = 1] = \frac{1}{6},$$

$$\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = \frac{1}{6} - \left(\frac{1}{6}\right)^2 = \frac{5}{36},$$

$$\mathrm{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] = 0 - \left(\frac{1}{6}\right)^2 = -\frac{1}{36}.$$

for $i \neq j$

Approximating the multinomial distribution for large n

By the LLN, as $n \to \infty$,

$$\left(\frac{Y_1}{n}, \dots, \frac{Y_6}{n}\right) \to \left(\frac{1}{6}, \dots, \frac{1}{6}\right)$$

in probability. By the CLT, as $n \to \infty$,

$$\sqrt{n}\left(\frac{Y_1}{n} - \frac{1}{6}, \dots, \frac{Y_6}{n} - \frac{1}{6}\right) \to \mathcal{N}(0, \Sigma)$$

in distribution, where

$$\Sigma = \begin{pmatrix} \frac{5}{36} & -\frac{1}{36} & \cdots & -\frac{1}{36} \\ -\frac{1}{36} & \frac{5}{36} & \cdots & -\frac{1}{36} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{36} & -\frac{1}{36} & \cdots & \frac{5}{36} \end{pmatrix} \in \mathbb{R}^{6 \times 6}.$$

(The negative values of Σ_{ij} for $i \neq j$ mean Y_i and Y_j are, as expected, slightly anti-correlated.)

Continuous mapping

The LLN and CLT can be used as building blocks to understand other statistics, via the **Continuous Mapping Theorem**:

Theorem

If $T_n \to c$ in probability, then $g(T_n) \to g(c)$ in probability for any continuous function g.

If $T_n \to Z$ in distribution, then $g(T_n) \to g(Z)$ in distribution for any continuous function g.

(These hold in both the univariate and multivariate settings.)

Is your friend cheating you in dice?

Recall

$$nT_n = n\left(\frac{Y_1}{n} - \frac{1}{6}\right)^2 + \ldots + n\left(\frac{Y_6}{n} - \frac{1}{6}\right)^2.$$

The function $g(x_1, \ldots, x_6) = x_1^2 + \ldots + x_6^2$ is continuous, so

$$nT_n \to Z_1^2 + \ldots + Z_6^2.$$

in distribution, where $(Z_1, \ldots, Z_6) \sim \mathcal{N}(0, \Sigma)$.

Hence, when n is large, the distribution of T_n is approximately that of $\frac{1}{n}(Z_1^2 + \ldots + Z_6^2)$.

Is your friend cheating you in dice?

Recall

$$nT_n = n\left(\frac{Y_1}{n} - \frac{1}{6}\right)^2 + \ldots + n\left(\frac{Y_6}{n} - \frac{1}{6}\right)^2.$$

The function $g(x_1, \ldots, x_6) = x_1^2 + \ldots + x_6^2$ is continuous, so

$$nT_n \to Z_1^2 + \ldots + Z_6^2.$$

in distribution, where $(Z_1, \ldots, Z_6) \sim \mathcal{N}(0, \Sigma)$.

Hence, when n is large, the distribution of T_n is approximately that of $\frac{1}{n}(Z_1^2 + \ldots + Z_6^2)$.

Finally, what is the distribution of $Z_1^2 + \ldots + Z_6^2$?

Is your friend cheating you in dice?

Using bilinearity of covariance, it is easy to show that if

$$W_1, \ldots, W_6 \overset{IID}{\sim} \mathcal{N}(0,1),$$

then

$$\frac{1}{\sqrt{6}}(W_1 - \bar{W}, \ldots, W_6 - \bar{W}) \sim \mathcal{N}(0, \Sigma).$$

(Here $\bar{W} = \frac{1}{6}(W_1 + \ldots + W_6)$.)

Is your friend cheating you in dice?

Using bilinearity of covariance, it is easy to show that if

$$W_1, \ldots, W_6 \overset{IID}{\sim} \mathcal{N}(0, 1),$$

then

$$\frac{1}{\sqrt{6}}(W_1 - \bar{W}, \ldots, W_6 - \bar{W}) \sim \mathcal{N}(0, \Sigma).$$

(Here $\bar{W} = \frac{1}{6}(W_1 + \ldots + W_6)$.)

So $Z_1^2 + \ldots + Z_6^2$ has the same distribution as

$$\frac{1}{6}\left((W_1 - \bar{W})^2 + \ldots + (W_6 - \bar{W})^2\right).$$

This is the sample variance of 6 IID standard normals, which we will show next week has distribution $\frac{1}{6}\chi_5^2$.

Conclusion: T_n has approximate distribution $\frac{1}{6n}\chi_5^2$.

Is your friend cheating you in dice?

Here's our simulated histogram of T_n, overlaid with the (appropriately rescaled) PDF of the $\frac{1}{6n}\chi^2_5$ distribution: