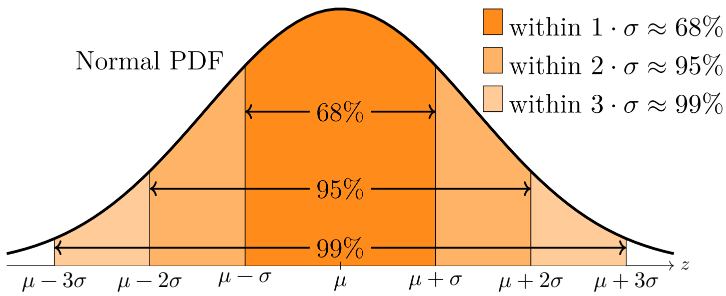
Class 6b in-class problems, 18.05, Spring 2022

Concept questions

Concept question 1. Normal distributions

X has normal distribution, standard deviation σ.



(a) *$P(-\sigma < X - \mu < \sigma)$ is approximately*

(i) 0.025 (ii) 0.16 (iii) 0.68 (iv) 0.84 (v) 0.95

(b) *$P(X > \mu + 2\sigma)$ is approximately*

(i) 0.025 (ii) 0.16 (iii) 0.68 (iv) 0.84 (v) 0.95

Solution: (a) Correct answer is (iii). The rule of thumb says the probability that X is within one standard deviation of the mean is 0.68.

(b) Correct answer is (i). This question for the probability in the right tail, beyond 2 standard deviations above the mean. The rule of thumb is that about 95% of the probability is within 2σ of the mean. So about 5% is outside of that. Since this is split symmetrically between two tails, the probability in the right tail is approximately 0.025.

Board questions

Problem 1. Standardization

Suppose X is a random variable with mean μ and standard deviation σ. Let Z be the standardization of X.

(a) *Give the formula for Z in terms of X, μ and σ.*

(b) *Use the algebraic properties of mean and variance to show Z has mean 0 and standard deviation 1.*

Solution: (a) $Z = \dfrac{X - \mu}{\sigma}$.

(b) The problem asks us to verify that $E[Z] = 0$ and $\text{Var}(Z) = 1$.

We use the properties

$$E[aX + b] = aE[X] + b = a\mu + b$$
$$\text{Var}(aX + b) = a^2\text{Var}(X) = a^2\sigma^2.$$

In the following, don't forget that $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$.

$$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}E[X-\mu] = \frac{1}{\sigma}(E[X]-\mu) = 0.$$
$$\text{Var}(Z) = \text{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\text{Var}(X-\mu) = \frac{1}{\sigma^2}\text{Var}(X) = \frac{1}{\sigma^2}\cdot\sigma^2 = 1.$$

Problem 2. CLT

(a) *Carefully write the statement of the central limit theorem.*

(b) *To head the newly formed US Dept. of Statistics, suppose that 50% of the population supports the team of Alessandre, Gabriel, Sarah and So Hee, 25% support Jen and 25% support Jerry.*

A poll asks 400 random people who they support. What is the probability that at least 55% of those polled prefer the team?

(c) *What is the probability that less than 20% of those polled prefer Jen?*

Solution: (b) Let $\overline{X}$ be the fraction polled who support the team. So $\overline{X}$ is the average of 400 Bernoulli(0.5) random variables. That is, let $X_i = 1$ if the ith person polled prefers the team and 0 if not, so $\overline{X}$ = average of the X_i.

The question asks for the probability $\overline{X} > 0.55$.

Each X_i has $\mu = 0.5$ and $\sigma^2 = 0.25$. So, $E[\overline{X}] = 0.5$ and $\sigma^2_{\overline{X}} = 0.25/400$ or $\sigma_{\overline{X}} = 1/40 = 0.025$.

Because $\overline{X}$ is the average of 400 Bernoulli(0.5) variables, the CLT says it is approximately normal and standardizing gives

$$\frac{\overline{X}-0.5}{0.025} \approx Z$$

So,

$$P(\overline{X} > 0.55) \approx P(Z > 2) \approx 0.025.$$

(c) Let $\bar{J}$ be the fraction polled who support Jen. The question asks for the probability that $\bar{J} < 0.2$.

Similar to part (b), $\bar{J}$ is the average of 400 Bernoulli(0.25) random variables. So,

$E[\bar{J}] = 0.25$ and $\sigma_S^2 = (0.25)(0.75)/400 \Rightarrow \sigma_S = \sqrt{3}/80$.

So, $\dfrac{\bar{J}-0.25}{\sqrt{3}/80} \approx Z$. Thus,

$$P(\bar{J} < 0.2) \approx P(Z < -4/\sqrt{3}) = \texttt{pnorm}(-4/\texttt{sqrt}(3), 0, 1) \approx 0.0105$$

Problem 3. Sampling from the standard normal distribution

How would you approximate a single random sample from a standard normal distribution using 9 rolls of a ten-sided die?

Note: *$\mu = 5.5$ and $\sigma^2 = 8.25$ for a single roll of a 10-sided die.*

Hint: *CLT is about averages.*

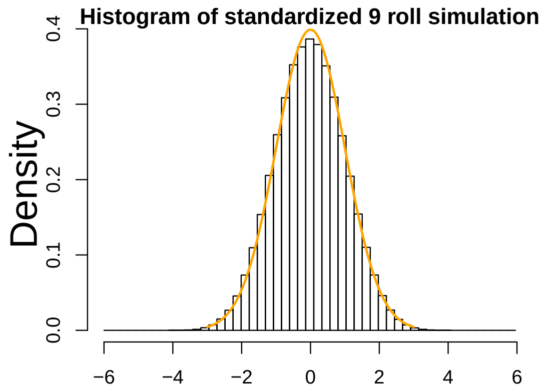
Solution: The average of 9 rolls is a sample from the average of 9 independent random variables. The CLT says this average is approximately normal with $\mu = 5.5$ and $\sigma = \sqrt{8.25/9} = 0.957$

If $\overline{x}$ is the average of 9 rolls then standardizing we get

$$z = \frac{\overline{x} - 5.5}{0.957}$$

is (approximately) one sample from $\mathrm{N}(0, 1)$.

So, to approximate a standard normal, we would roll 9 times and compute z.



Standard normal is shown in orange.

$\overline{X}$ = average of nine rolls: $\mu = 5.5$, $\sigma = \sqrt{8.25/9}$.

Standarized statistic: $Z = \dfrac{\overline{X} - \mu}{\sigma} \approx N(0, 1)$.

Extra problems

Bonus problem

An accountant rounds to the nearest dollar. We'll assume the error in rounding is uniform on [-0.5, 0.5]. Estimate the probability that the total error in 300 entries is more than \$5.

Solution: Let X_j be the error in the j^{th} entry, so, $X_j \sim U(-0.5, 0.5)$.

We have $E[X_j] = 0$ and $\mathrm{Var}(X_j) = 1/12$.

The total error $S = X_1 + ... + X_{300}$ has $E[S] = 0$, $\mathrm{Var}(S) = 300/12 = 25$, and $\sigma_S = 5$.

Standardizing we get, by the CLT, $S/5$ is approximately standard normal. That is, $S/5 \approx Z$.

So, $P(S < -5 \text{ or } S > 5) \approx P(Z < -1 \text{ or } Z > 1) \approx \boxed{0.32}$.