When one files a tax return, one has to add some n dollar amounts. If one rounds off each amount to the nearest whole dollar, what overall error should one expect?

Probabilistic model: \( D_i = \text{exact amounts}, \quad \text{(e.g.}$178.43$) \)
\( Z_i = \text{roundoff errors} \sim \text{Unif} \left[-\frac{1}{2}, \frac{1}{2}\right] \text{ independent} \quad (-0.43) \)

- Each amount is rounded off to \( X_i = D_i + Z_i \quad (\$178) \)
- Total is \( \sum_{i=1}^{n} X_i \)

\[
\text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var} (X_i) \quad \text{(by independence)}
\]
\[
= \sum_{i=1}^{n} \text{Var} (Z_i) = \frac{n}{12} \quad \text{(since Var} (Z_i) = \frac{1}{12}, \text{ see Uniform distribution })
\]

St dev. of \( \sum_{i=1}^{n} X_i \) is \( \sqrt{\frac{n}{12}} \), Ans.

\[\sqrt{\frac{n}{12}} \leq \frac{n}{2} \quad \text{(worst-case error)} \quad \text{E.g.,} \quad n = 50, \quad \sqrt{\frac{50}{12}} = 2.0 \quad \text{(while} \frac{n}{2} = 25)\]

- Application: quantization (probabilistic models)

**Example 10.14 (Investment)** Dr. Caprio has invested money in three uncorrelated assets: 25% in the first one, 43% in the second one, and 32% in the third one. The means of the annual rate of returns for these assets, respectively, are 10%, 15%, and 13%. Their standard deviations are 8%, 12%, and 10%, respectively. Find the mean and standard deviation of the annual rate of return for Dr. Caprio’s total investment.

Solution: Let \( r \) be the annual rate of return for Dr. Caprio’s total investment. Let \( X_1, X_2, \) and \( X_3 \) be the annual rate of returns for the first, second, and third assets, respectively. By Example 4.25,
\[ X = 0.25X_1 + 0.43X_2 + 0.32X_3. \]

Thus
\[ E(X) = 0.25E(X_1) + 0.43E(X_2) + 0.32E(X_3) \]
\[ = (0.25)(0.10) + (0.43)(0.15) + (0.32)(0.13) = 0.1311. \]

Since the assets are uncorrelated, by (10.11),
\[ \text{Var} (X) = (0.25)^2 \text{Var}(X_1) + (0.43)^2 \text{Var}(X_2) + (0.32)^2 \text{Var}(X_3) \]
\[ = (0.25)^2(0.08)^2 + (0.43)^2(0.12)^2 + (0.32)^2(0.10)^2 = 0.004087. \]

Therefore, \( \sigma_X = \sqrt{0.004087} = 0.064. \) Hence Dr. Caprio should expect an annual rate of return of 13.11% with standard deviation of 6.4%. Note that Dr. Caprio has reduced the standard deviation of his investments considerably by diversifying his investment; that is, by not putting all of his eggs in one basket.
Application to statistics: sample mean, variance

Example: Interested in the distribution of income of a population
- Survey: take a random sample of \( n \) independently chosen people's incomes.
- Infer the distribution of income of the population from the distribution of income in the sample.

Does this method always work?
What is the error?
How large should the sample \( n \) be?

Probabilistic (statistical) model:

\( r.v. \quad X = \text{income of a randomly chosen person (from population)} \)

Want to compute:
- \( \mu = E[X] \), the "population mean" (of income)
- \( \sigma^2 = \text{Var}(X) \), the "population variance" (of income)

Sample: \( \{X_1, X_2, \ldots, X_n\} \), independent r.v's representing the incomes of the people selected for the sample.
- All \( X_i \) have the same distribution as \( X \).
- "\( X_i \) are independent identically distributed (iid) copies of \( X \)."

1. Estimating the population mean.

**Def.** Sample mean: \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \)

\( \bar{X} \) is also a r.v. (varies depending on who we included in the sample).

Our hope is that

\[ \mu \approx \bar{X} \]

population mean, \( \mu \), unknown

sample mean, \( \bar{X} \), an "estimator" of \( \mu \)
\[ E(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \]

\textbf{Proof}

\[ E(X) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu \]

\[ \text{Var}(\bar{X}) = \frac{1}{n^2} \text{Var}(\sum_{i=1}^{n} X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{n \sigma^2}{n^2} = \frac{\sigma^2}{n} \quad \text{(by independence)} \]

\[ \text{QED.} \]

\textbf{Remarks:}
1) \( E(\bar{X}) = \mu \) means that \( \bar{X} \) is an \textit{unbiased} estimator of \( \mu \).
2) \( \sigma_{\bar{X}} = \sqrt{\text{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}} \) is the standard error of such estimation.

\textbf{Note:} \( \sigma_{\bar{X}} \to 0 \) as \( n \to \infty \)

\( \Rightarrow \) \( \bar{X} \) becomes better estimate as the sample increases.

2. Estimating the population variance

\textbf{Def.} \textbf{Sample variance:} \( s^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \).

Again, \( s^2 \) is a random variable, and our hope is that \( \text{Var}(s^2) = \sigma^2 \).

\textbf{THM} \quad E[s^2] = \sigma^2 \quad \text{(i.e. \( s^2 \) is an unbiased estimator of \( \sigma^2 \)).}

\textbf{Proof.} We want to replace \( \bar{X} \) by \( \mu \) in the def. of \( s^2 \):

\[ (n-1)s^2 = \sum_{i=1}^{n} [(X_i - \mu) - (\bar{X} - \mu)]^2 = \sum_{i=1}^{n} (X_i - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^{n} (X_i - \mu) + n(\bar{X} - \mu)^2 \]

\( = \sum_{i=1}^{n} (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \)

Now take expectations of both sides:

\[ (n-1) E[s^2] = \sum_{i=1}^{n} E[(X_i - \mu)^2] - n E[(\bar{X} - \mu)^2] \]

\[ E[(X_i - \mu)^2] = \sigma^2 \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad \text{(by Prop. above)} \]

\[ = n \sigma^2 - \sigma^2 = (n-1) \sigma^2. \quad \text{QED.} \]
\[ X = \text{amount of pain-relieving medicine}, \]
\[ Y = \text{level of pain after taking the medicine} \]

Model:
\[ \begin{align*}
Y &= b - mX + Z \\
\text{\textit{ initial lead effect of pain medicine}} \\
\text{\textit{ random fluctuations, indep. of X.}} \\
X &\sim N(\mu_x, \sigma_x^2); \quad Z \sim N(\mu_z, \sigma_z^2) \quad \text{independent.}
\end{align*} \]

Compute \( p_{X,Y} \).

\[
\text{Cov}(X, Y) = \frac{\text{Cov}(X, b)}{\sigma_X} - m \frac{\text{Cov}(X, X)}{\text{Var}(X)} + \frac{\text{Cov}(X, Z)}{\text{Var}(Z)} \quad \text{by independence}
\]

\[ = -m \text{ Var}(X) \]

\[
\text{Var}(Y) = \text{Var}(-mX + Z) = m^2 \text{Var}(X) + \text{Var}(Z)
\]

\[
p_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{-m \sigma_x^2}{\sigma_X \sqrt{m^2 \sigma_x^2 + \sigma_z^2}} = \frac{-1}{\sqrt{1 + \left( \frac{\sigma_z^2}{m \sigma_x^2} \right)^2}}
\]

- \( \sigma_z = 0 \) (no fluctuations) \( \Rightarrow \) \( p_{X,Y} = -1 \) (strongest)
- \( \sigma_z \to \infty \) (overwhelming fluctuations) \( \Rightarrow \) \( p_{X,Y} = 0 \) (no pattern).