Ex 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 = exact amounts, (e.g. \$178.43)

Z_i = roundoff errors $\sim \text{Unif}[-\frac{1}{2}, \frac{1}{2}]$ independent (−\$0.43)

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

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) \quad \text{(by independence)}$$

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

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

$\sqrt{\frac{n}{12}} \ll \frac{n}{2}$ (worst-case error) E.g. $n = 50$, $\sqrt{\frac{n}{12}} = \$2.0$ (while $\frac{n}{2} = \$25$)

- Application quantization (probabilistic models)

04/06 Lec. 36

★ **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 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 (Ex 4a)

Example [Interested in the distribution of income of AA population
Survey: take a random sample of n indep. people's incomes.
Infer the distr of income of population from the distr. 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. X = income of a randomly chosen person (from population)

Want to compute:

$\mu = E[X]$, the "population mean" (of income)

$\sigma^2 = Var(X)$, the "population variance" (of income)

Sample: $\{X_1, X_2, \dots, X_n\}$, 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, unknown

sample mean, known; an "estimator" of μ

Prop $\boxed{E[\bar{X}] = \mu, \quad Var(\bar{X}) = \frac{\sigma^2}{n}}$

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

$Var(\bar{X}) = \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{Var(X_i)}_{=\sigma^2}$ (by independence)

$= \frac{\sigma^2}{n}$

QED.

Remarks: 1) $E[\bar{X}] = \mu$ means that $\bar{X}$ is an "unbiased" estimator of μ.

2) $\sigma_{\bar{X}} = \sqrt{Var(\bar{X})} = \frac{\sigma}{\sqrt{n}}$ is the standard error of such estimation.

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

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

Again, s^2 is a random variable, and our hope is that

$$s^2 \approx \sigma^2$$

THM $\boxed{E[s^2] = \sigma^2}$ (i.e. s^2 is an unbiased estimator of σ^2)

Proof. We want to replace $\bar{X}$ by μ in the def. of s^2:

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

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

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

Now take expectations of both sides:

$$(n-1)\, E[s^2] = \sum_{i=1}^{n} \underbrace{E[(X_i - \mu)^2]}_{E[(X-\mu)^2] = \sigma^2} - n \cdot \underbrace{E[(\bar{X} - \mu)^2]}_{Var(\bar{X}) = \frac{\sigma^2}{n} \text{ (by Prop. above)}}$$

$$= n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

QED.

Ex | X = amount of pain-relieving medicine,

Y = level of pain after taking the medicine

Model:

- $Y = \underset{\substack{\uparrow \\ \text{initial level} \\ \text{of pain}}}{b} - \underbrace{mX}_{\text{effect of medicine}} + \underbrace{Z}_{\text{random fluctuations, indep. of } X}$ ("regression of Y on X")

- $X \sim N(\mu_X, \sigma_X^2)$; $Z \sim N(\mu_Z, \sigma_Z^2)$ independent.

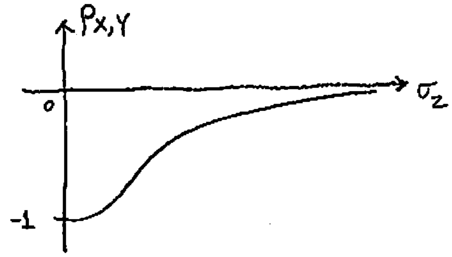
Compute $\rho_{X,Y}$.

$$Cov(X,Y) = \underbrace{Cov(X,b)}_{\substack{\| \\ 0\ (b=\text{const})}} - m \underbrace{Cov(X,X)}_{\substack{\| \\ Var(X)}} + \underbrace{Cov(X,Z)}_{\substack{\| \\ 0 \text{ by independence}}}$$

$$= -m\, Var(X)$$

$$Var(Y) = Var(-mX + Z) = m^2 Var(X) + Var(Z)$$

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X\ \sigma_Y} = \frac{-m\,\sigma_X^2}{\sigma_X\sqrt{m^2\sigma_X^2 + \sigma_Z^2}} = -\frac{1}{\sqrt{1 + \left(\frac{\sigma_Z}{m\sigma_X}\right)^2}}.$$



$\sigma_Z = 0$ (no fluctuations) $\Rightarrow \rho_{X,Y} = -1$ (strongest)

$\sigma_Z \to \infty$ (overwhelming fluctuations)

$\Rightarrow \rho_{X,Y} = 0$ (no pattern).