Deviations from the Mean

Reading

- Malik Magdon-Ismail. Discrete Mathematics and Computing.
  - Chapter 21

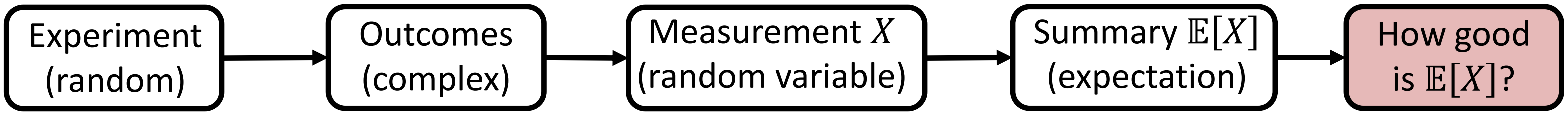
Overview

Rensselaer

- How well does the expected value (mean) summarize a random variable?
- Variance
- Variance of a sum
- Law of large numbers
  - The 3-σ rule

Probability for Analyzing a Random Experiment

- An experiment typically has the following stages



  - We want to know how likely is it that we got an unlikely outcome?
  - How much did we learn from that experiment?
- **Experiment.** Roll n dice and compute X, the average of the rolls

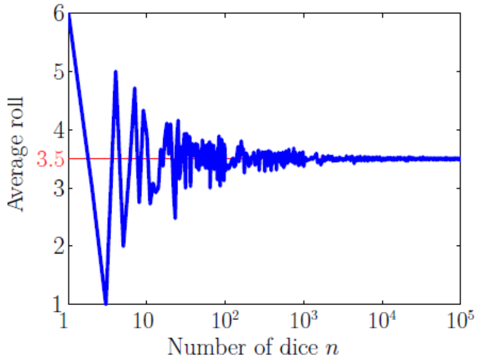
$$\mathbb{E}[average] = \mathbb{E}\left[\frac{1}{n} \times sum\right] = \frac{1}{n}\mathbb{E}[sum]$$

$$= \frac{1}{n} \times n \times 3.5 = 3.5$$

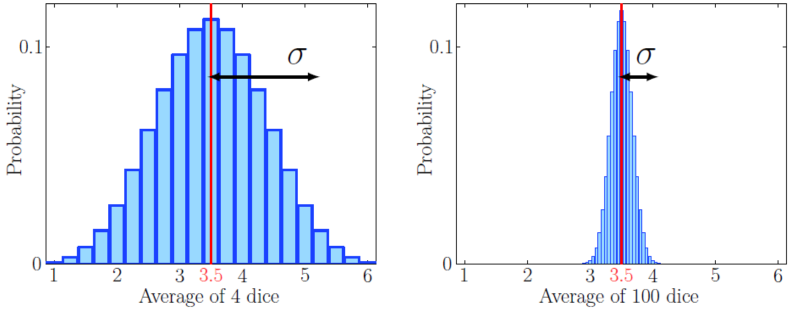
- The expected average is the same as a single die roll average!
  - Expectation doesn’t tell us anything about the spread of the values we’ll see

Average of n Dice

- Here's what happens when you roll more dice (and calculate the average)!
  - Exciting stuff!



  - Huh, looks like the average varies very little for large n
- This is not surprising if you calculate the PDF



  - The variable σ measures "the spread" of the PDF
  - Notice that the PDF is much more compact for larger n

Variance: Size of the Deviations from the Mean

Rensselaer

- As usual, let's start with X, the sum of 2 dice

$$\mathbb{E}[X] = 7$$

  - Denote the mean by $\mu(X) = \mathbb{E}[X]$
  - Let $\Delta = X - \mu$, a random variable that measures the deviation from the mean

X	2	3	4	5	6	7	8	9	10	11	12
Δ	-5	-4	-3	-2	-1	0	1	2	3	4	5
$P_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- **Variance,** σ^2, is the expected value of the squared deviations,

$$\sigma^2 = \mathbb{E}\left[\Delta^2\right] = \mathbb{E}\left[(X - \mu)^2\right] = \mathbb{E}\left[(X - E[X])^2\right]$$

  - Why not just take the expected value of Δ?

$$\mathbb{E}[\Delta] = 0$$

- In the dice example,

$$\sigma^2 = \mathbb{E}\left[\Delta^2\right] = \frac{1}{36} \times 25 + \frac{2}{36} \times 16 + \frac{3}{36} \times 9 + \cdots + \frac{1}{36} \times 25 \approx 5.83$$

Variance, cont'd

Rensselaer

- **Variance,** σ^2, is the expected value of the squared deviations,

$$\sigma^2 = \mathbb{E}\left[\Delta^2\right] = \mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}\left[(X-E[X])^2\right]$$

- In the dice example,

$$\sigma^2 = \mathbb{E}\left[\Delta^2\right] = \frac{1}{36}\times 25 + \frac{2}{36}\times 16 + \frac{3}{36}\times 9 + \cdots + \frac{1}{36}\times 25 \approx 5.83$$

- **Standard Deviation,** σ, is the square-root of the variance

$$\sigma = \sqrt{\mathbb{E}[\Delta^2]} = \sqrt{\mathbb{E}[(X-\mu)^2]} = \sqrt{\mathbb{E}[(X-E[X])^2]}$$

$$\sigma = \sqrt{5.83} \approx 2.52$$

- The standard deviation is typically used to construct confidence intervals around your mean
  - You will often hear statisticians report the mean $\pm$ 1 standard deviation

sum of two dice rolls = 7 $\pm$ 2.52

- **Practice.** Exercise 21.2.

Variance is a Measure of Risk

Rensselaer

- Suppose I give you 2 possible games
- Game 1.

$$X_1 = \begin{cases} win\ \$2 \quad probability = \frac{2}{3} \\ lose\ \$1 \quad probability = \frac{1}{3} \end{cases}$$

$$\mathbb{E}[X_1] = \$1$$

- Game 2.

$$X_1 = \begin{cases} win\ \$102 \quad probability = \frac{2}{3} \\ lose\ \$201 \quad probability = \frac{1}{3} \end{cases}$$

$$\mathbb{E}[X_1] = \$1$$

- Which one do you prefer?

Variance is a Measure of Risk, cont'd

Rensselaer

- Suppose I give you 2 possible games
  - Game 1.

$$X_1 = \begin{cases} win\ \$2 \quad probability = \frac{2}{3} \\ lose\ \$1 \quad probability = \frac{1}{3} \end{cases}$$

$$\mathbb{E}[X_1] = \$1$$

  - Game 2.

$$X_1 = \begin{cases} win\ \$102 \quad probability = \frac{2}{3} \\ lose\ \$201 \quad probability = \frac{1}{3} \end{cases}$$

$$\mathbb{E}[X_1] = \$1$$

- Let's calculate the variances:

$$\sigma^2(X_1) = \frac{2}{3}(2-1)^2 + \frac{1}{3}(-1-1)^2 = 2$$

$$\sigma^2(X_2) = \frac{2}{3}(102-1)^2 + \frac{1}{3}(-201-1)^2 \approx 2 \times 10^4$$

- So,

$$X_1 = 1 \pm \sqrt{2}$$

$$X_2 = 1 \pm 100\sqrt{2}$$

- For a small expected profit you might risk a small loss (Game 1), not a huge loss (Game 2).

A More Convenient Formula for Variance

Rensselaer

- Let's play around with the variance definition

$$\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right]$$
$$= \mathbb{E}\left[X^2 - 2X\mu + \mu^2\right] \quad \textbf{[expand the square]}$$
$$= \mathbb{E}\left[X^2\right] - 2\mu\mathbb{E}[X] + \mu^2 \quad \textbf{[Linearity of expectation]}$$
$$= \mathbb{E}\left[X^2\right] - 2\mu^2 + \mu^2 \quad [\mathbb{E}[X] = \mu]$$
$$= \mathbb{E}\left[X^2\right] - \mu^2$$

- **Variance**: $\sigma^2 = \mathbb{E}\left[X^2\right] - \mu^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$
- Sum of 2 dice:

$$\mathbb{E}\left[X^2\right] = \sum_{x=2}^{12} P_X(x)x^2$$
$$= \frac{1}{36}2^2 + \frac{2}{36}3^2 + \cdots + \frac{1}{36}12^2 \approx 54.83$$

- Since $\mu = 7$, $\sigma^2 = 54.83 - 7^2 = 5.83$
- **Theorem.** Variance ≥ 0, which means $\mathbb{E}\left[X^2\right] \geq \mathbb{E}[X]^2$ for any random variable X

Variance of Uniform Random Variable

- Let X be drawn from a uniform distribution. We saw earlier that $\mathbb{E}[X] = \frac{1}{2}(n+1)$
- Let's now calculate $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \frac{1}{n}\left(1^2 + \cdots + n^2\right) =$$
$$= \frac{1}{n} \times \frac{n}{6}(n+1)(2n+1) =$$
$$= \frac{1}{6}(n+1)(2n+1)$$

- So, what is the variance?

$$\sigma^2(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2 =$$
$$= \frac{1}{12}(n+1)(4n+2-3n-3)$$
$$= \frac{1}{12}\left(n^2 - 1\right)$$

Variance of Bernoulli Random Variable

- We saw earlier that $\mathbb{E}[X] = p$
- What is $\mathbb{E}\left[X^2\right]$?

$$\mathbb{E}\left[X^2\right] = p \times 1^2 + (1-p) \times 0^2 = p$$

- So, the variance is:

$$\sigma^2(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = p - p^2 = p(1-p)$$

Linearity of Variance

Rensselaer

- We know expectation is a linear operator, but how about variance?
- Let's look at an example first
- Let X be a Bernoulli variable and let $Y = a + X$, where a is a constant
  - What is the PDF of Y?

$$Y = a + 1, \text{ with probability } p$$

$$Y = a, \text{ with probability } 1 - p$$

- What is $\mathbb{E}[Y]$?

$$\mathbb{E}[Y] = p \times (a + 1) + (1 - p) \times a = p + a = a + \mathbb{E}[X]$$

  - Linear as expected
- How about the deviations Δ_Y

$$\Delta_Y = 1 - p, \text{ with probability } p$$

$$\Delta_Y = -p, \text{ with probability } 1 - p$$

  - Deviations don't depend on a!
- Therefore, $\sigma^2(Y) = \sigma^2(X)$

Linearity of Variance, cont'd

- *Theorem*. Let $Y = a + bX$. Then,

$$\sigma^2(Y) = b^2\sigma^2(X)$$

Variance of a Sum

Rensselaer

- Let X_1 and X_2 be any two random variables
- What is the variance $\sigma^2(X)$, where $X = X_1 + X_2$?
- Let's calculate the relevant quantities:

$$\mathbb{E}[X]^2 = (\mathbb{E}[X_1 + X_2])^2 = (\mathbb{E}[X_1] + \mathbb{E}[X_2])^2 \qquad \textbf{[linearity of expectation]}$$
$$= \mathbb{E}[X_1]^2 + 2\mathbb{E}[X_1]\mathbb{E}[X_2] + \mathbb{E}[X_2]^2$$
$$\mathbb{E}[X^2] = \mathbb{E}[(X_1 + X_2)^2] = \mathbb{E}[X_1^2 + 2X_1X_2 + X_2^2]$$
$$= \mathbb{E}[X_1^2] + 2\mathbb{E}[X_1X_2] + \mathbb{E}[X_2^2] \qquad \textbf{[linearity of expectation]}$$

- Then, the variance is

$$\sigma^2(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 =$$
$$= \left(\mathbb{E}[X_1^2] + 2\mathbb{E}[X_1X_2] + \mathbb{E}[X_2^2]\right) - \left(\mathbb{E}[X_1]^2 + 2\mathbb{E}[X_1]\mathbb{E}[X_2] + \mathbb{E}[X_2]^2\right)$$
$$= \left(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2\right) + \left(\mathbb{E}[X_2^2] - \mathbb{E}[X_2]^2\right) + 2(\mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2])$$
$$= \sigma^2(X_1) + \sigma^2(X_2) + 2(\mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2])$$

  - Note that last term is 0 if X_1 and X_2 are independent

- **Variance of a Sum.** For *independent* random variables, the variance of the sum is a sum of the variances.

Variance of a Sum, cont'd

Rensselaer

- **Practice.** Compute the variance of 1 die roll. Compute the variance of the sum of n dice rolls.
- **Example**. The Variance of the Binomial (sum of *independent* Bernoullis)
  - Let $X = X_1 + \cdots + X_n$ (sum of independent Bernoullis)
  - Let $\sigma(X_i) = p(1-p)$. Show that

$$\begin{aligned}\sigma^2(Binomial) &= \sigma^2(X_1) + \cdots + \sigma^2(X_n) \\ &= p(1-p) + \cdots + p(1-p) = np(1-p)\end{aligned}$$

3-σ Rule: $X = \mu(X) \pm 3\sigma(X)$

- **3-σ Rule**. For *any* random variable X, the chances are at least (about) 90% that

$$\mu - 3\sigma < X < \mu + 3\sigma \qquad \text{or} \qquad X = \mu(X) \pm 3\sigma$$

  - Allows us to judge the spread of a distribution by knowing σ
  - Also allows us to judge the quality of our current measurement

- *Lemma [Markov Inequality]*. For a positive random variable X,

$$\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[X]}{\alpha}$$

- *Proof*.

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x \geq 0} x P_X(x) \\ &\geq \sum_{x \geq \alpha} x P_X(x) \\ &\geq \sum_{x \geq \alpha} \alpha P_X(x) \\ &= \alpha \mathbb{P}[X \geq \alpha] \end{aligned}$$

- Allows to bound the tail end of the distribution

3-σ Rule: $X = \mu(X) \pm 3\sigma(X)$

Rensselaer

- **3-σ Rule**. For *any* random variable X, the chances are at least (about) 90% that

$$\mu - 3\sigma < X < \mu + 3\sigma \qquad \text{or} \qquad X = \mu(X) \pm 3\sigma$$

  - Allows us to judge the spread of a distribution by knowing σ
  - Also allows us to judge the quality of our current measurement
- *Lemma [Markov Inequality]*. For a positive random variable X,

$$\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[X]}{\alpha}$$

- *Lemma [Chebyshev Inequality]*.

$$\mathbb{P}[|X - \mu| \geq t\sigma] \leq \frac{1}{t^2}$$

- *Proof*.

$$\mathbb{P}[|X - \mu| \geq t\sigma] = \mathbb{P}\left[|X - \mu|^2 \geq t^2\sigma^2\right]$$

$$\leq \frac{\mathbb{E}\left[|X-\mu|^2\right]}{t^2\sigma^2} \qquad \text{[Markov Inequality]}$$

$$= \frac{\sigma^2}{t^2\sigma^2} = \frac{1}{t^2}$$

- To get the 3-σ rule, use Chebyshev's Inequality with $t = 3$

Law of Large Numbers

- We know that the expected average of a sum is unchanged with n
- E.g., expectation of the average of n dice:

$$\mathbb{E}[average] = \mathbb{E}\left[\frac{1}{n} \times sum\right] = \frac{1}{n}\mathbb{E}[sum] = \frac{1}{n} \times n \times 3.5 = 3.5$$

- But the variance gets smaller and smaller:

$$\sigma^2(average) = \sigma^2\left(\frac{1}{n} \times sum\right) = \frac{1}{n^2} \times \sigma^2(sum)$$
$$= \frac{1}{n^2} \times n \times \sigma^2(one\ die) = \frac{1}{n} \times \sigma^2(one\ die)$$

  - As we increase n, the variance goes down to 0
- This is another way of saying that the sample average gets closer to $\mathbb{E}[one\ die]$ as n gets large
  - Law of Large Numbers!