Section 3.4

Variance

Ex) $X \sim Unif[-1000, 1000], \quad Y \sim Unif[-1,1]$

Both have same mean (i.e., expected value), $EX = EY = 0$. But X "varies much more", i.e., is usually much farther from its mean.

We need a statistic to capture this.

suggestion: $\mathbb{E}\left(X - \underbrace{\mathbb{E}X}_{\text{const}}\right)^2$

Variance

$\mathbb{E}X$

Def: Let X be a r.v. with mean μ. The variance of X is

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

Notation: $\sigma^2 = \sigma^2(X) = Var(X) =$ "variance of X"

$\sigma := \sqrt{Var(X)} =$ "standard deviation of X"

Note: $\sigma = \sqrt{\mathbb{E}(X-\mu)^2} \neq \mathbb{E}(X-\mu)$

Ex) $X = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases}$ $\quad \mu = 0$

RHS: $\mathbb{E}X - 0 = 0$

LHS: $\sqrt{\mathbb{E}X^2} = \sqrt{\mathbb{E}1} = \sqrt{1} = 1$

$\sigma(aX) = |a|\,\sigma(X)$

Calculating $Var(X)$

Set $g(X) = (X - \mu)^2$ so $Var(X) = \mathbb{E}g(X)$.

Now use formula for $\mathbb{E}g(X)$:

Calculating variance:

Discrete case: $Var(X) = \sum_k \overbrace{(k - \mu)^2}^{g(k)} p(k)$

Continuous case: $Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$

Recall: $\mu \coloneqq \mathbb{E}X$

Ex) $X \sim Unif[-1000, 1000]$, $\quad Y \sim Unif[-1,1]$

In both cases, $\mu = \underline{0}$

$$Var(X) = \mathbb{E}\,\underline{(X-\mu)^2} = \mathbb{E}\,\underline{X^2} = \int_{-1000}^{1000} X^2 \cdot \frac{1}{2000} = \frac{1}{2000} \cdot \frac{x^3}{3}\Bigg|_{-1000}^{1000}$$

$$= \boxed{\frac{2(1000)^3}{3 \cdot 2000}}$$

$$Var(Y) = \underline{\int_{-1}^{1} x^2 \cdot \frac{1}{2}\,dx} = \frac{x^3}{3 \cdot 2}\Bigg|_{-1}^{1} = \boxed{\frac{2 \cdot 1^3}{3 \cdot 2}}$$

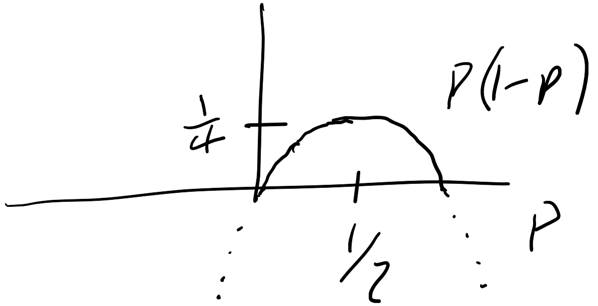
Ex) $X \sim Bern(p)$, $Var(X) = ?$

$$\mu = \mathbb{E}X = p$$

$$Var(X) = \mathbb{E}(X-\mu)^2 = \mathbb{E}(X-p)^2$$

$$= \underset{\substack{\uparrow \\ p(0)}}{(1-p)} \cdot (-p)^2 + \underset{\substack{\uparrow \\ p(1)}}{p} \cdot (1-p)^2$$

$$= (1-p)\,p\,\cancel{(p+1-p)} = \boxed{p(1-p)}$$



Facts about variance, expectation

1. $Var(X) = 0 \quad \Leftrightarrow \quad$ X is a constant
2. $\mathbb{E}[X_1 + X_2 + \ldots + X_n] = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}\,X_n$

   always, as long as RHS exists

3. $Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$

   if $X_1, \ldots, X_n$ are indep. & RHS exists

4. $\mathbb{E}[a\,X + b] =$ $a\,\mathbb{E}[X] + b$
5. $Var(aX + b) =$ $a^2\,Var(X)$
6. $\boxed{Var(X) = \mathbb{E}(X-\mu)^2 = \mathbb{E}X^2 - \mu^2}$

Property 6, method to calculate variance

$$Var(X) = \mathbb{E}(X-\mu)^2 = \mathbb{E}X^2 - \mu^2$$

Proof (of second equality): $\mathbb{E}(X-\mu)^2 = \mathbb{E}\left[X^2 - 2\mu \cdot X + \mu^2\right]$

$= \mathbb{E}X^2 - 2\mu \cdot \mathbb{E}X + \mathbb{E}\mu^2$

$= \mathbb{E}X^2 - 2\mu \cdot \mu + \mu^2 = \mathbb{E}X^2 - \mu^2$ ✓

Q) You choose from two games to play:

1. Flip four coins, you win $1 for each head, or

X = winnings $\qquad X \sim Bin(\underset{n}{4}, \underset{p}{\tfrac{1}{2}})$

2. Flip one coin, if it is heads, you win $4.

Y = winnings

Which is better?

$\mathbb{E}X = np = 4 \cdot \frac{1}{2} = 2.$

$\mathbb{E}Y = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 4 = 2.$

Which has lower variance?

$Var(X) = Var(X_1 + X_2 + X_3 + X_4)$

Bern($\frac{1}{2}$)

$= Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4)$

$= 4 \cdot Var(Bern(\frac{1}{2}))$

$= 4 \cdot \frac{1}{2} \cdot \frac{1}{2} = \boxed{1}.$

$Var(Y) = Var(4 \cdot Bern(\frac{1}{2}))$

$= 16 \cdot \frac{1}{2} \cdot \frac{1}{2} = \boxed{4}$

Ex) $\underbrace{X \sim Bin(n,p).}$ $Var(X) = Var(X_1 + X_2 + \dots + X_n)$

Bern(p)

$= Var(X_1) + \dots + Var(X_n) = \boxed{n \cdot p(1-p)}$

Ex) $X \sim Geom(p).$ $Var(X) = \underbrace{\mathbb{E}X^2}_{A} - \underbrace{(\mathbb{E}X)^2}_{\left(\frac{1}{p}\right)^2}$

Set $A = \mathbb{E}X^2 = \sum_{k=1}^{\infty} k^2 \cdot (1-p)^{k-1} \cdot p$

$= \cancel{1^2 \cdot (1-p)^0} \cdot p + \sum_{k=2}^{\infty} k^2 (1-p)^{k-1} p$

$$= p + \sum_{k=1}^{\infty} (k+1)^2 (1-p)^k \, p$$

$$= p + (1-p)\sum_{k=1}^{\infty} (k^2 + 2k + 1)(1-p)^{k-1} \, p$$

$$= p + (1-p)\left[\underbrace{\sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p}_{A} + \underbrace{2 \underbrace{\sum_{k=1}^{\infty} k(1-p)^{k-1} p}_{EX}}_{\frac{2}{p}} + \underbrace{\sum (1-p)^{k-1} p}_{\text{sum of pmf} = 1}\right]$$

$$\Rightarrow \quad A = p + (1-p)\left(A + \frac{2}{p} + 1\right)$$

$$A - (1-p)A = \frac{p^2 + 2(1-p) + (1-p)p}{p}$$

$$pA = \frac{\cancel{p^2} + 2 - 2p + p - \cancel{p^2}}{p} = A = \frac{2-p}{p^2}$$

$$Var(X) = A - \frac{1}{p^2}$$

$$= \boxed{\frac{1-p}{p^2}}$$