# Probability I - Lecture 2

Daniel Yu

September 12, 2024

# Contents

# 1 Expectation

**Definition 1.** If $X$ is a discrete random variable, then the expectation (**expected value**) is:

$$E[X] = \sum_{k \in range(X)} k \cdot P[X = k].$$

aka a space average. If $X$ is a continuous random variable with pdf $f_x\,(t)$,

$$E[X] = \int_{-\infty}^{\infty} t \cdot f_x(t) dt.$$

**Remark.** Examples

- $E[Bernoulli(\theta)] = 0 \cdot (1-p) + 1 \cdot p = p$
- Let $Y$ be the outcome of a fair 3 sided die. $E[Y] = \frac{1+2+3+4}{4} = 2.5$.
- M = maxiumum of two fair 4-sided dice. Let $M$ be random variable representing the maximum of two fair dice rolls (independent). The $\Omega = \{(1,1), \ldots\}$. Also, $P$ is uniform for dice rolls. $E[M] = 1 \cdot P[M = 1] + 2 \cdot P[M = 2] + 3 \cdot P[M = 3] + 4 \cdot P[M = 4] = 1 \cdot \frac{1}{16} + 2\frac{3}{16} + 3\frac{5}{16} + 4\frac{7}{16}$
- $R \sim Geo(p)$ geometric distribution. Define $P[R = k] = p(1-p)^k$ and $k \in \{1, 2, 3, \ldots\}$.
- $x \sim \exp(\lambda)$ i.e. $f_x(t) = \lambda e^{-\lambda t}$ when $t \geq 0$ and 0 otherwise.

**Theorem 1.** If $X$ and $Y$ are two random varirables defined on the same space and $a, b \in \mathbb{R}$. Then,

$$E[aX + bY] = aE[X] + bE[Y].$$

*Proof.* Discrete case

$$
\begin{aligned}
E[Z] &= \sum_{k \in range(z)} k \cdot P[Z = k] \\
&= \sum_{r \in range(X), t \in range(y)} (ar + bt) P[Z = ar + bt] \\
&= \sum_{r,t} (ar + bt) \cdot P[X = r, Y = t] \text{ — Note r,t is shorthand} \\
&= \sum_r \sum_t ar P[X = r, Y = t] + \sum_r \sum_t bt \cdot P[X = r, Y = t] \\
&= \sum_r ar \cdot (\sum_t P[X = r, Y = t) + \sum_t bt \sum_r P[X = r, Y = t] \\
&= a \sum_r r \cdot P[X = r] + b \sum_t t \cdot P[Y = t] \\
&= aE[X] + bE[Y].
\end{aligned}
$$

□

This is useful! For example, back to card dealing, assuming we agin deal 3 cards from 52 cards. Observe $A = X+Y+Z$, we can use $E[A] = E[X]+E[Y]+E[Z] = \frac{3}{13}$

# 2 Higher Moments

$$E[x^k] = \text{ kth moment of X.}$$

## 2.1 Variance

**Definition 2.** The second moment known as the **variance**: $E[x^2]$:

$$var\,(x) = E[x^2] - E[x]^2.$$

**Lemma 1.**

$$var\,(x) = E[(X - E[X])^2].$$

**Remark.** The two methods of computation are equivalent but different. For example:

$$E[Ber(p)^2] = p, E[Ber(p)] = p^2 \to var(Ber(p)) = p - p^2.$$

Then,

$$Y = Ber(p) - E[Ber(p)] = Ber(p) - p.$$

where $Y$ is the distribution:

$$P[Y = p - 1] = pP[Y = -p] = 1 - p.$$

and the distribution of $Y^2$:

$$P[Y^2 = (1-p)^2] = p^2 P[Y^2 = p^2] = 1 - p.$$

so when we compute we get the same answer:

$$\begin{aligned} var(Ber(p)) &= E[(Ber(p) - E[Ber(p)])] = E[Y^2] \\ &= (1-p)^2 \cdot p + p^2 \cdot (1-p) \\ &= p(1-p) \\ &= p - p^2. \end{aligned}$$

*Proof.* Let $\mu = E[X]$

$$\begin{aligned} E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \\ &= E[X^2] - E[X]^2. \end{aligned}$$

□

**Remark.** Variance can be thought of as a measure of how **random** a R.V. is.

**Proposition 1.** If $var(x) = 0 \Leftrightarrow P[x = a] = 1$ for some $a$ i.e. $X$ is not a random variable at all!

*Proof.* Assume $var(x) = 0$. Then this means $E[(X-\mu)^2] = 0$. □

**Remark.** Example
$X = \{1, 2, 3, 4\}$ with P uniform. $E[X] = \frac{5}{2}$.

$$E[X^2] = 1 \cdot P[X^2 = 1] + 4 \cdot P[X^2 = 4] + 9 \cdot P[X^2 = 9] + 16 \cdot P[X^2 = 16].$$

$$= \frac{1 + 4 + 9 + 16}{4} = \frac{15}{2}.$$

So,

$$var(X) = \frac{15}{2} - \frac{5}{2}^2 = \frac{5}{4}.$$

**Remark.** Example
$U \sim uniform(0, 2)$ continuous. $E[U] = \int_0^2 t \cdot \frac{1}{2}$ since uniform probability would be .5 across interval from 0 to 2.

$$var(u) = E[(U-1)^2] = \int_0^2 (t-1)^2 \cdot .5dt.$$

$$= \frac{1}{2} \int_0^2 t^2 - 2t + 1dt.$$

$$= \frac{1}{2}[\frac{t}{3}^3 - t^2 + t]_0^2.$$
$$= \frac{1}{3}.$$

**Proposition 2.** 1. Variance is not preserved under scalar multiplication

$$\begin{aligned} var(aX) &= E[(aX)^2] - E[aX]^2 \\ &= E[a^2 X_2] - (aE[x])^2 \\ &= a^2(E[X^2] - E[X]^2) \\ &= a^2 \cdot var(X). \end{aligned}$$

2. variance is preserved under additivity

$$var(X + b) = var(X).$$

3. Variance is nonlinear

$$var(X + Y) \neq var(X) + (Y) \text{ in general.}$$

**Definition 3.** We have a special term to measure the relationship of between the variance sof 2 random variables known as the **covariance**

$$cov(X, Y) = E[XY] - E[X]E[Y].$$

The matrix covariance formula for $X, Y$ matrices is: $Cov(X, Y) = E[(X - E[X])(Y - E[Y])^T]$

**Lemma 2.**

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

*Proof.*

$$\begin{aligned} E[(X+Y)^2] &= E[X^2] + 2E[XY] + E[Y^2] \\ E[X+Y]^2 &= E[X]^2 + 2E[X]E[Y] + E[Y]^2 \\ \text{Var}(X+Y) &= (E[X^2] + 2E[XY] + E[Y^2]) - (E[X]^2 + 2E[X]E[Y] + E[Y]^2) \\ &= E[X^2] - E[X]^2 + 2(E[XY] - E[X]E[Y]) + E[Y^2] - E[Y]^2 \\ &= \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y) \end{aligned}$$

□

Recall the example from the previous lecture

**Note.** For example, consider the discrete random variables $x, y, z$ over $\Omega = \{1,2,3,4\}$ on a uniform probability measure:

$$X = \begin{cases} 1, w \in \{1,2\} \\ 0, w \in \{3,4\} \end{cases} .$$

$$Y = \begin{cases} 1, w \in \{3,4\} \\ 0, w \in \{1,2\} \end{cases} .$$

$$z = \begin{cases} 1, w \in \{1,3\} \\ 0, w \in \{2,4\} \end{cases} .$$

What is the variance of $X + Y$?

$$var(X+Y) = Var(x) + Var(y) + 2Cov(X,Y) = \frac{1}{4} + \frac{1}{4} + 2Cov(X,Y).$$

$$Cov(X,Y) = E[XY] - E[X]E[Y] = E[XY] \cdot \frac{1}{4} = -\frac{1}{4}.$$

$$var(X+Y) = \frac{1}{2} - 2 \cdot \frac{1}{4} = 0!$$

**Definition 4.** If $Cov(X,Y) = 0$ we say $X, Y$ are uncorrelated

# 3 Independence

**Definition 5.** Let $\Omega$ be the outcome space and $P$ the probability measure. $A, B \subseteq \Omega$ are events. $A, B$ are **independent** if $P[A \cap B] = P[A] \cdot P[B]$. We say $X, Y$ random variables on $\Omega$ are independent if $\{X = a\}$ and $\{Y = b\}$ are independent events for every a,b in range of X, Y respectively.