Probability and Random Process

COSE382

Expectation

Definition 4.1.1(Expectation of a discrete r.v.). The *expected value* (also called the *expectation* or *mean*) of a discrete r.v. X is defined by

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j) = \sum_x \underbrace{x}_{value} \underbrace{p_X(x)}_{\text{PMF at}x},$$

a weighted average of the possible values that X can take on.

Examples

1. $X \sim \text{DUnif}(\{1, 2, 3, 4, 5, 6\})$

$$E(X) = \frac{1}{6}(1 + 2 + \cdots + 6) = 3.5,$$

2. Let $X \sim \text{Bern(p)}$ and $q = 1 - p$. Then

$$E(X) = 1p + 0q = p,$$

Note that $E(X)$ depends only on the *distribution* of X.

Linearity of Expectation

Theorem 4.2.1(Linearity of expectation.). For any r.v.s X, Y and any constant c,

$$
\begin{aligned}
E(X+Y) &= E(X)+E(Y), \\
E(cX) &= cE(X).
\end{aligned}
$$

Proof: will be presented later (after *joint distribution* is introduced)

Example 4.2.2(Binomial expectation). For $X \sim \mathrm{Bin}(n,p)$,

$$E(X) = np$$

Proof in two ways

1) Since X is the sum of n i.i.d. Bernoulli r.v.s I_j

$$X = I_1 + \cdots + I_n,$$

where each I_j has expectation $E(I_j) = 1p + 0q = p$ By linearity,

$$E(X) = E(I_1) + \cdots + E(I_n) = np.$$

2) From $k\binom{n}{k} = n\binom{n-1}{k-1}$,

$$\begin{aligned} E(X) = \sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k} &= n\sum_{k=0}^{n}\binom{n-1}{k-1}p^k q^{n-k} = np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}q^{n-k} \\ &= np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j q^{n-1-j} = np \end{aligned}$$

Definition 4.3.1(Geometric R.V.).

$X \sim \text{Geom}(p)$: In a sequence of i.i.d. Bernoulli trials with success probability p, X is a number of *failures* before the first success. X is called *Geometric random variable* or *Geometric distribution*

Theorem 4.3.2(Geometric PMF). If $X \sim \text{Geom}(p)$, then the PMF of X is

$$P(X = k) = (1-p)^k p \quad \text{for } k = 0, 1, 2, \cdots$$

This is a valid PMF because

$$\sum_{k=0}^{\infty} (1-p)^k p = p \sum_{k=0}^{\infty} (1-p)^k = p \cdot \frac{1}{1-(1-p)} = 1.$$

Example:

A newlywed couple plans to have children and will continue until the first girl. What is the probability that the couple have a boy and a girl ?

Solution:

$$p_X(1) = (1/2)^1 \, (1/2)^1 = 0.25$$

Definition 4.3.4(First Success distribution).

$Y \sim \mathrm{FS}(p)$: In a sequence of i.i.d. Bernoulli trials with success probability p, Y is the number of *trials* to have the first success.

- PMF of $Y \sim \mathrm{FS}(p)$

$$Y(k) = (1-p)^{k-1}p, \quad k = 1, 2, \cdots$$

Example:

Products produced by a machine has a 3% defective rate. What is the probability that the first defective occurs in the fifth item inspected?

Solution:

$$p_X(5) = (0.03)^1 (0.97)^4 \approx 0.02655878$$

- Relation between $\mathrm{Geom}(p)$ and $\mathrm{FS}(p)$
  - If $X \sim \mathrm{Geom(p)}$ then $X + 1 \sim \mathrm{FS}(p)$

$$P(X+1=k) = P(X=k-1) = (1-p)^{k-1}p, \quad k = 1, 2, \cdots$$

  - If $Y \sim \mathrm{FS}(p)$ then $Y - 1 \sim \mathrm{Geom}(p)$

$$P(Y-1=k) = P(Y=k+1) = (1-p)^{k}p, \quad k = 0, 1, \cdots$$

Example 4.3.5(Geometric Expectation). Let $X \sim \text{Geom}(p)$, then

$$E(X) = \sum_{k=0}^{\infty} k(1-p)^k p = \frac{1-p}{p}.$$

Proof: Consider $\sum_{k=0}^{\infty}(1-p)^k = \frac{1}{p}$. Differentiating both sides with respect to p,

$$\sum_{k=0}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2}. \quad \text{Hence,}$$

$$E(X) = \sum_{k=0}^{\infty} k(1-p)^k p = p(1-p)\sum_{k=0}^{\infty} k(1-p)^{k-1} = p(1-p)\frac{1}{p^2} = \frac{1-p}{p}.$$

Example 4.3.6(First Success expectation). For $Y \sim \text{FS}(p)$, $Y = X + 1$ $(X \sim \text{Geom}(p))$,

$$E(Y) = E(X+1) = \frac{1-p}{p} + 1 = \frac{1}{p}.$$

Example 4.3.11 (Coupon collector). There are n types of toys. Assume that each time you get a toy, it is equally likely to be any of the n types. What is the expected number of toys needed until you have a complete set?

Solution: Let N be the number of toys needed; we want to find $E(N)$. Our strategy will be to break up N into a sum of simpler r.v.s so that we can apply linearity.

$$N = N_1 + N_2 + \cdots + N_n,$$

where N_i is the additional numbers of toys until the ith new toy type after $i-1$ toy types. Then, $N_1 = 1$, $N_2 \sim \mathrm{FS}(\frac{n-1}{n})$ and

$$N_i \sim \mathrm{FS}(\frac{n-i+1}{n}).$$

$$\begin{aligned} E(N) &= E(N_1) + E(N_2) + \cdots + E(N_n) = 1 + \frac{n}{n-1} + \cdots + n \\ &= n(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{1}) = n\sum_{i=1}^{n} \frac{1}{i}. \end{aligned}$$

For large n, this is very close to $n(\log n + 0.577)$.

Example 4.3.13 (St. Petersburg paradox). If you flip a fair coin n times to see Head for the first time, you win the game and will receive $\$2^n$.

1) What is the expected number of trials to win ?
2) What is the expected money you will get ?

Solution: Let N be the number of trials to win the game.

1) From $N \sim FS(1/2)$,

$$E(N) = 2$$

2) Let X be the prize when the game ends at N trials. $X = 2^N$.

$$E(X) = \sum_{k=1}^{\infty} 2^k P(N = k) = \sum_{k=1}^{\infty} 2^k \frac{1}{2^k} = \infty$$

Here, we have $2^{E(N)} = 4 \neq \infty$, illustrating the danger of confusing $E(g(X))$ with $g(E(X))$ when g is not linear.

Law of unconscious statistician

Theorem 4.5.1 (LOTUS). If X is a discrete r.v. and $g : \mathbb{R} \to \mathbb{R}$, then for $Y = g(X)$

$$E(Y) = E(g(X)) = \sum_x g(x)P(X = x),$$

where the sum is taken over all possible values of X.

Proof.

$$\begin{aligned} E(Y) &= \sum_y yP(Y = y) \\ &= \sum_{\{x|g(x)=y\}} g(x)P(X = x) \\ &= \sum_x g(x)P(X = x). \end{aligned}$$

Variance

Definition 4.6.1 (Variance and standard deviation). The *variance* of an r.v. X is

$$\mathrm{Var}(X) := E(X - EX)^2.$$

The square root of the variance is called the *standard deviation (SD)*:

$$\mathrm{SD}(X) := \sqrt{Var(X)}$$

Recall that when we write $E(X - EX)^2$, we mean the expectation of the random variable $(X - EX)^2$, not $(E(X - EX))^2$ (which is 0 by linearity).

Theorem 4.6.2 For any r.v. X,

$$\mathrm{Var}(X) = E(X^2) - (EX)^2.$$

Proof. Let $\mu = EX$. Expand $(X - \mu)^2$ and use linearity:

$$\mathrm{Var}(X) = E(X - \mu)^2 = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2.$$

Properties of Variance

- $\text{Var}(X + c) = \text{Var}(X)$
- $\text{Var}(cX) = c^2\text{Var}(X)$
- If X and Y are independent, then $\text{Var}(X + Y) =$Var(X)+Var(Y)
- $\text{Var}(X) \geq 0$

(Variance is not linear). Unlike expectation, variance is *not* linear.

- The constant comes out *squared* in $\text{Var}(cX) = c^2\ \text{Var}(X)$
- The variance of the sum of r.v.s may or may not be the sum of their variances. For example, if $X = Y$,

$$\text{Var}(X + Y) = \text{Var}(2X) = 4\text{Var}(X) \neq 2\text{Var}(X)$$

Example 4.6.4 (Variance of Geometric R.V). Let $X \sim \mathrm{Geom}(p)$.

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X = k) = \sum_{k=0}^{\infty} k^2 p(1-p)^k = \sum_{k=1}^{\infty} k^2 p(1-p)^k.$$

$$\sum_{k=1}^{\infty} k(1-p)^k = \frac{1-p}{p^2}, \quad \frac{d}{dt} \sum_{k=1}^{\infty} k(1-p)^k = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} = \frac{2-p}{p^3}.$$

Hence,

$$E(X^2) = \sum_{k=1}^{\infty} k^2 p(1-p)^k = p(1-p)\frac{2-p}{p^3} = \frac{(1-p)(2-p)}{p^2}.$$

Finally,

$$\mathrm{Var}(X) = E(X^2) - (EX)^2 = \frac{(1-p)(2-p)}{p^2} - \left(\frac{1-p}{p}\right)^2 = \frac{1-p}{p^2}$$

– For $Y \sim \mathrm{FS}(p)$, $\mathrm{Var}(Y) = \mathrm{Var}(X) = \dfrac{1-p}{p^2}$.

Example 4.6.5(Binomial variance). For $X \sim \mathrm{Bin}(n, p)$, $X = Y_1 + Y_2 + \cdots + Y_n$, where $Y_j \sim \mathrm{Bern}(p)$. Each Y_j has variance

$$\mathrm{Var}(Y_j) = E(Y_j^2) - (E(Y_j))^2 = p - p^2 = p(1-p)$$

Since the $\{Y_i\}$ are independent, we have

$$\mathrm{Var}(X) = \mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_n) = np(1-p).$$

Poisson random variable

Definition 4.7.1 (Poisson distribution). $X \sim \text{Pois}(\lambda)$: X has the *Poisson distribution* with parameter λ, where $\lambda > 0$, if the PMF of X is

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

- This is a valid PMF because of the Taylor series

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}.$$

- In practice,
  - λ is interpreted as a *rate* of occurrence of rare events of interest during a time interval
  - $P(X = k)$ tells the probability that k events occurs during the time interval.

Example (Poisson distribution). Customers of an Internet service provider initiate new accounts at the average rate of 10 accounts per day.

- What is the probability that more than 2 new accounts will be initiated today ?
- What is the probability that more than 4 accounts will be initiated within 2 days ?

Solution. Let X be the number of initiated new accounts in a day. $X \sim \text{Pois}(\lambda)$, where $\lambda = 10$ accounts per day. $p_X(k) = e^{-10}\dfrac{10^k}{k!}$

$$P(X > 2) = 1 - p_X(0) - p_X(1) - p_X(2) = 1 - e^{-10}(1 + 10 + 10^2/2) = 0.9995$$

Let Y be the number of initiated new accounts in two days. $Y \sim \text{Pois}(2\lambda)$

$$\begin{aligned} P(X > 4) &= 1 - p_y(0) - p_Y(1) - p_Y(2) - p_Y(3) - p_Y(4) \\ &= 1 - e^{-20}(1 + 20 + 20^2/2 + 20^3/6 + 20^4/24) \approx 1 \end{aligned}$$

Example 4.7.2 (Poisson expectation and variance). Let $X \sim \text{Pois}(\lambda)$.

$$\begin{aligned} E(X) &= e^{-\lambda}\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda}e^{\lambda} = \lambda. \end{aligned}$$

Consider

$$\begin{aligned} \frac{d}{d\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} &= \sum_{k=1}^{\infty} k\frac{\lambda^{k-1}}{k!} = e^{\lambda} \\ \frac{d}{d\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!} &= \sum_{k=1}^{\infty} k^2\frac{\lambda^{k-1}}{k!} = \frac{d}{d\lambda}\lambda e^{\lambda} = e^{\lambda} + \lambda e^{\lambda} = e^{\lambda}(1+\lambda), \end{aligned}$$

Finally,

$$\begin{aligned} E(X^2) &= e^{-\lambda}\sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda}\lambda(1+\lambda) = \lambda(1+\lambda), \\ \text{Var}(X) &= E(X^2) - (EX)^2 = \lambda(1+\lambda) - \lambda^2 = \lambda. \end{aligned}$$

Note that $E(X) = \text{Var}(X) = \lambda$

Connection between Poisson and Binomial

Theorem 4.8.3 (Poisson approximation to Binomial). If $X \sim \text{Bin}(n, p)$ and we let $n \to \infty$ and $p \to 0$ such that $\lambda = np$ remains fixed, then the PMF of X converges to the Pois(λ) PMF.

Proof. Since $\lambda = np$, $p = \lambda/n$

$$
\begin{aligned}
P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} \frac{n(n-1)...(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}
\end{aligned}
$$

Letting $n \to \infty$ with k fixed,

$$
\frac{n(n-1)...(n-k-+1)}{n^k} \to 1, \quad \left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}, \quad \left(1 - \frac{\lambda}{n}\right)^{-k} \to 1,
$$

Hence,

$$
\lim_{n \to \infty} P(X = k) \to \frac{e^{-\lambda} \lambda^k}{k!},
$$

Theorem 4.8.1 (Sum of independent Poissons). If $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2)$, and X is independent of Y, then $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$

Proof.

Method 1: Consider two Binomial r.v.s $X_b \sim \text{Bin}(n_1, p)$ and $Y_b \sim \text{Bin}(n_2, p)$ where $n_1 p = \lambda_1$ and $n_2 p = \lambda_2$. $X_b + Y_b \sim \text{Bin}(n_1 + n_2, p)$ and as $n_1, n_2 \to \infty$ with $n_1 p = \lambda_1$, $n_2 p = \lambda_2$ remain fixed, $X_b + Y_b \to X + Y$ and $X_b + Y_b \to \text{Pois}(\lambda_1 + \lambda_2)$

Method 2: Alternatively, to get the PMF of $X + Y$, condition on X and use the law of total probability:

$$\begin{aligned} P(X+Y=k) &= \sum_{j=0}^{k} P(X+Y=k|X=j)P(X=j) = \sum_{j=0}^{k} P(Y=k-j)P(X=j) \\ &= \sum_{j=0}^{k} \frac{e^{-\lambda_2}\lambda_2^{k-j}}{(k-j)!}\frac{e^{-\lambda_1}\lambda_1^{j}}{j!} = \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{j=0}^{k}\binom{k}{j}\lambda_1^j\lambda_2^{k-j} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^k}{k!}. \end{aligned}$$

Theorem 4.8.2 (Poisson given a sum of Poissons). If $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$, and X is independent of Y, then the conditional distribution of X given $X+Y=n$ is $\mathrm{Bin}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$.

Proof.

$$
\begin{aligned}
P(X=k|X+Y=n) &= \frac{P(X+Y=n|X=k)P(X=k)}{P(X+Y=n)} \\
&= \frac{P(Y=n-k)P(X=k)}{P(X+Y=n)} \\
&= \frac{\left(\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\right)\left(\frac{e^{-\lambda_1}\lambda_1^{k}}{k!}\right)}{\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}} \\
&= \binom{n}{k}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}
\end{aligned}
$$

which is the $\mathrm{Bin}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$ PMF, as desired.