Discrete Random Variables

Dr. Edmund Lam

Department of Electrical and Electronic Engineering
The University of Hong Kong

ELEC2844: Probabilistic Systems Analysis

(Second Semester, 2021–22)

https://www.eee.hku.hk/~elec2844

Random variables

Deterministic world:

- Numbers (arithmetic): $3 + 4 = 7$
- Variables (algebra): $x = 7$
- Functions of variables: $g(x) = x^2 \log_2 x$

Probabilistic world:

- Numbers: $\mathbb{P}(\text{head}) = \frac{1}{2}$
- Variables: $X =$ sum of two dice rolls
- Functions of variables: $g(X) = X^2 \log_2 X$

We are interested in questions such as $\mathbb{P}(X = 7)$, the distribution of X, properties of such distribution (e.g., mean), $\mathbb{P}(g(X) = 192)$, the distribution of $g(X)$, etc.

Random variables

Random variable (r.v.) X:

- takes on numerical values
  - We may need to *assign* numerical values to each outcome
- convention is to use capital letter
- discrete random variable means the set of values it can take is either finite or countably infinite
- the probability of X being equal to each possible value in its range forms a probability mass function (PMF)

$$\boxed{\text{PMF} = \text{“ideal” histogram}} \tag{1}$$

- a function of X leads to another PMF

1. Probability mass function
2. Expectation, mean, variance
3. Common discrete random variables
4. Joint PMF, conditioning, independence

Notation

A r.v. X can take on different numerical values. If x is any real number within the range of this r.v., we write

$$\boxed{p_X(x) = \mathbb{P}(X = x)} \tag{2}$$

By definition, $\sum_x p_X(x) = 1$. Also, $p_X(x)$ must be nonnegative.

Example: Let $X =$ number of heads in two tosses of a fair coin.

$$p_X(x) = \begin{cases} \frac{1}{4} & x = 0 \\ \frac{1}{2} & x = 1 \\ \frac{1}{4} & x = 2 \\ 0 & \text{otherwise.} \end{cases}$$

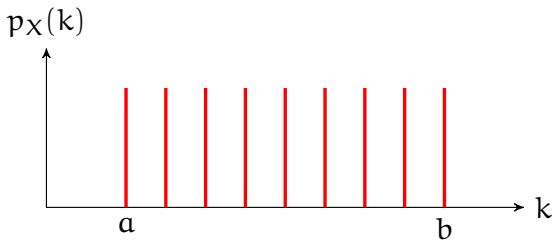
Archetype

1. $p_X(x)$ where $X =$ dice roll
2. $p_X(x)$ where we have a biased coin with probability p for head, and you obtain $X = 1$ if that happens
3. $p_X(x)$ where we have a biased coin with probability p for head, and you obtain $X = 1$ if that happens, which you repeat n times
4. $p_X(x)$ where we have a biased coin with probability p for head, and $X =$ number of tosses until head appears
5. $p_X(x)$ where $X =$ number of cars having accidents on a given day with large number of cars, low probability for each car to be in an accident, but a fixed "rate"

(1) Discrete uniform random variable

Discrete uniform random variable over a range $[a, b]$ has PMF

$$\boxed{p_X(k) = \begin{cases} \dfrac{1}{b-a+1} & k = a, a+1, \ldots, b \\ 0 & \text{otherwise.} \end{cases}} \tag{3}$$



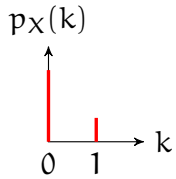
(2) Bernoulli random variable

Bernoulli random variable is useful for various binary state events: true/false, yes/no, 0/1, present/absent, etc. It has two possible values

$$X = \begin{cases} 1 & \text{head} \\ 0 & \text{tail} \end{cases}$$

Bernoulli PMF (coin toss with $\mathbb{P}(\text{head}) = p$) is

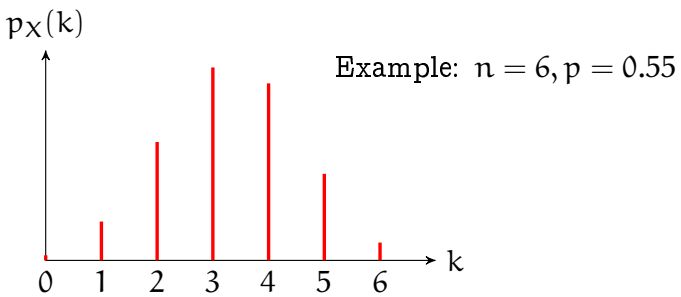
$$\boxed{p_X(k) = \begin{cases} p & k = 1 \\ 1-p & k = 0 \end{cases}} \tag{4}$$



(3) Binomial random variable

Binomial random variable is the sum of n Bernoulli trials with probability p, with PMF

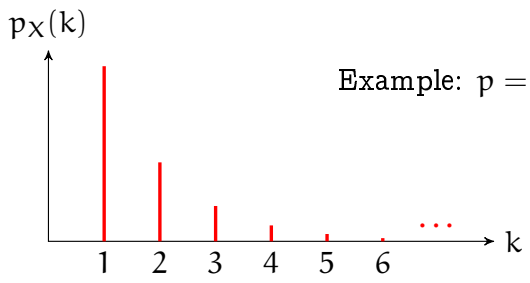
$$\boxed{p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n} \tag{5}$$



(4) Geometric random variable

Geometric random variable is the number of Bernoulli trials with probability p ($\mathbb{P}(\text{head}) = p$) before head turns up, with PMF

$$\boxed{p_X(k) = (1-p)^{k-1} p \qquad k = 1, 2, \ldots} \tag{6}$$



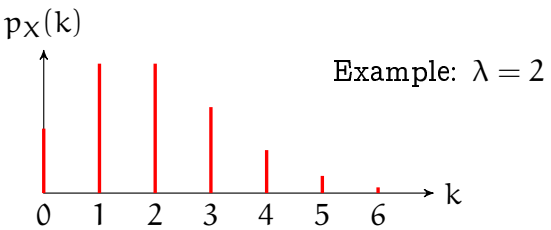
(5) Poisson random variable

Poisson random variable with parameter λ has PMF

$$\boxed{p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad k = 0, 1, \ldots} \tag{7}$$

It is a good approximation of binomial r.v. with $\lambda = np$, n is very large, and p very small, i.e.

$$\frac{e^{-\lambda}\lambda^k}{k!} \approx \binom{n}{k} p^k (1-p)^{n-k} \qquad k \ll n$$



Poisson random variable

Proof: For a binomial r.v. with parameters (n, p), and let $\lambda = np$,

$$\begin{aligned} p_X(k) &= \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{(np)^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \end{aligned}$$

We fix k but let $n \to \infty$, so for "small" $j = 1, \ldots, k$,

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

Therefore,

$$p_X(k) \to \frac{e^{-\lambda}\lambda^k}{k!}$$

Function of a random variable

If X is a r.v., and $Y = g(X)$ being a function of a r.v., then Y is also a r.v.

$$\boxed{p_Y(y) = \sum_{\{x \mid g(x)=y\}} p_X(x)} \tag{8}$$

Example: Suppose X is a r.v. such that

$$p_X(k) = \begin{cases} \frac{1}{9} & k = -4, -3, \ldots, 3, 4 \\ 0 & \text{otherwise.} \end{cases}$$

Find the PMF of (1) $Y = |X|$; (2) $Z = X^2$.

Example

Ans: $p_Y(0) = p_X(0) = \frac{1}{9}$, and then $p_Y(k) = p_X(-k) + p_X(k) = \frac{2}{9}$ for $k = 1, 2, 3, 4$. Therefore,

$$p_Y(k) = \begin{cases} \frac{2}{9} & k = 1, 2, 3, 4 \\ \frac{1}{9} & k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, $p_Z(0) = p_X(0) = \frac{1}{9}$, and then $p_Z(k^2) = p_X(-k) + p_X(k) = \frac{2}{9}$ for $k = 1, 2, 3, 4$. Therefore,

$$p_Z(k) = \begin{cases} \frac{2}{9} & k = 1, 4, 9, 16 \\ \frac{1}{9} & k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

1. Probability mass function
2. Expectation, mean, variance
3. Common discrete random variables
4. Joint PMF, conditioning, independence

Expectation

The expectation or expected value or mean of a r.v. X is

$$\boxed{\mathbb{E}(X) = \sum_x x\, p_X(x)} \tag{9}$$

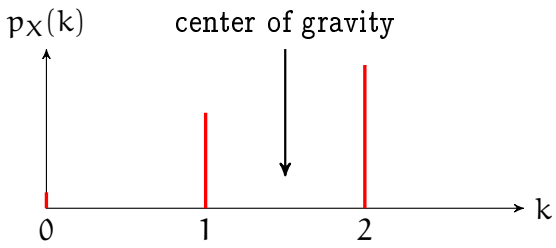
It is a very useful number as a "representative" value of X, and is also viewed as the "center of gravity."

Example

Consider two independent coin tosses with $\mathbb{P}(H) = \frac{3}{4}$, and let X be the number of heads. Find $p_X(k)$ and $\mathbb{E}(X)$.

Ans: There are only three possibilities: $p_X(k) = \begin{cases} \frac{1}{16} & k = 0 \\ \frac{3}{8} & k = 1 \\ \frac{9}{16} & k = 2 \end{cases}$

So the mean is $\mathbb{E}(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{9}{16} = \frac{3}{2}$



Moment, variance, standard deviation

The nth moment of r.v. X is given by $\mathbb{E}(X^n)$

$$\text{first moment} = \mathbb{E}(X) = \text{mean} \tag{10}$$

$$\text{second moment} = \mathbb{E}(X^2) \tag{11}$$

A related concept is the variance of a r.v.,

$$\mathrm{var}(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) \tag{12}$$

The standard deviation is the square root of variance,

$$\sigma_X = \sqrt{\mathrm{var}(X)} \tag{13}$$

Expectation for function of random variable

We can compute the expected value of $Y = g(X)$ by

$$\mathbb{E}(g(X)) = \sum_x g(x) p_X(x) \tag{14}$$

Proof:

$$\begin{aligned}
\mathbb{E}(g(X)) = \mathbb{E}(Y) = \sum_y y\, p_Y(y) &= \sum_y y \left(\sum_{\{x \mid g(x) = y\}} p_X(x) \right) \\
&= \sum_y \sum_{\{x \mid g(x) = y\}} y p_X(x) \\
&= \sum_y \sum_{\{x \mid g(x) = y\}} g(x) p_X(x) \\
&= \sum_x g(x) p_X(x)
\end{aligned}$$

Moment, variance

Moment and variance are both function of a random variable, so

$$\mathbb{E}(X^n) = \sum_x x^n p_X(x) \tag{15}$$

$$\mathrm{var}(X) = \sum_x \big(x - \mathbb{E}(X)\big)^2 p_X(x) \tag{16}$$

Moment, variance

If $Y = aX + b$, then

$$\boxed{\begin{aligned} \mathbb{E}(Y) &= a\mathbb{E}(X) + b \qquad (17)\\ \mathrm{var}(Y) &= a^2\mathrm{var}(X) \qquad (18)\end{aligned}}$$

Proof:

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_x (ax+b)p_X(x) = a\sum_x x\,p_X(x) + b\sum_x p_X(x) = a\mathbb{E}(X) + b\\
\mathrm{var}(Y) &= \sum_x \big(ax + b - \mathbb{E}(aX+b)\big)^2 p_X(x)\\
&= \sum_x \big(ax + b - a\mathbb{E}(X) - b\big)^2 p_X(x)\\
&= \sum_x a^2\big(x - \mathbb{E}(X)\big)^2 p_X(x) = a^2\mathrm{var}(X)
\end{aligned}$$

Variance

A very useful formula for the variance:

$$\boxed{\mathrm{var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2} \qquad (19)$$

Proof:

$$\begin{aligned}
\mathrm{var}(X) &= \sum_x \big(x - \mathbb{E}(X)\big)^2 p_X(x)\\
&= \sum_x \big(x^2 - 2x\mathbb{E}(X) + (\mathbb{E}(X))^2\big)p_X(x)\\
&= \sum_x x^2 p_X(x) - 2\mathbb{E}(X)\sum_x x\,p_X(x) + (\mathbb{E}(X))^2\sum_x p_X(x)\\
&= \mathbb{E}(X^2) - 2\big(\mathbb{E}(X)\big)^2 + \big(\mathbb{E}(X)\big)^2\\
&= \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2
\end{aligned}$$

Example

Consider two independent coin tosses with $\mathbb{P}(H) = \frac{3}{4}$, and let X be the number of heads. We already know $p_X(k)$ and $\mathbb{E}(X)$ from a previous calculation. Find $\mathrm{var}(X)$ using different formulas.

Method 1: Use Eq. (12). Let $Y = (X - \mathbb{E}(X))^2 = (X - \frac{3}{2})^2$. When $x = 0$ we have $y = \frac{9}{4}$; when $x = 1$ we have $y = \frac{1}{4}$; when $x = 2$ we have $y = \frac{1}{4}$. So,

$$p_Y(k) = \begin{cases} \frac{1}{16} & k = \frac{9}{4}\\ \frac{3}{8} + \frac{9}{16} = \frac{15}{16} & k = \frac{1}{4}\end{cases}$$

$\mathrm{var}(X) = \mathbb{E}(Y) = \frac{9}{4}\cdot\frac{1}{16} + \frac{1}{4}\cdot\frac{15}{16} = \frac{3}{8}$

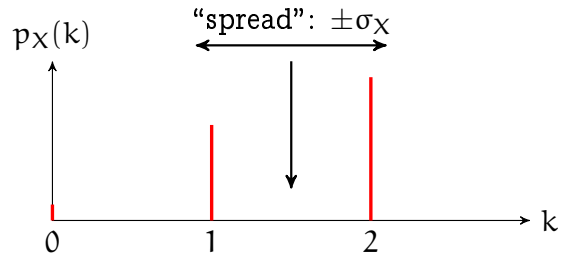
Example

Method 2: Use Eq. (16).

$$\mathrm{var}(X) = \sum_{x=0}^{2}(x - \tfrac{3}{2})^2 p_X(x) = \tfrac{9}{4}\cdot\tfrac{1}{16} + \tfrac{1}{4}\cdot\tfrac{3}{8} + \tfrac{1}{4}\cdot\tfrac{9}{16} = \tfrac{3}{8}$$

Method 3: Use Eq. (19). $\mathbb{E}(X^2) = 0^2\cdot\frac{1}{16} + 1^2\cdot\frac{3}{8} + 2^2\cdot\frac{9}{16} = \frac{21}{8}$. Therefore,

$$\mathrm{var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \tfrac{21}{8} - (\tfrac{3}{2})^2 = \tfrac{3}{8}$$



Example

We roll two dice. Let X be the absolute difference between the two numbers facing up. Find $p_X(k)$, $\mathbb{E}(X)$, and $\text{var}(X)$.

Ans: We note that k can take on integer values between 0 and 5.

- $\{1,6\}$, $\{6,1\}$: $k=5$ with prob. $p_X(5)=2/36$
- $\{1,5\}$, $\{2,6\}$, $\{6,2\}$, $\{5,1\}$: $k=4$ with prob. $p_X(4)=4/36$
- $\{1,4\}$, ...: $k=3$ with prob. $p_X(3)=6/36$
- $\{1,3\}$, ...: $k=2$ with prob. $p_X(2)=8/36$
- $\{1,2\}$, ...: $k=1$ with prob. $p_X(1)=10/36$
- $\{1,1\}$, ...: $k=0$ with prob. $p_X(0)=6/36$

Useful to check that the sum of these probabilities is 1.

Example

We apply the expectation formula directly:

$$\mathbb{E}(X) = \sum_{k=0}^{5} k p_X(k)$$
$$= (5\cdot 2 + 4\cdot 4 + 3\cdot 6 + 2\cdot 8 + 1\cdot 10)/36 = \frac{35}{18}$$

We also calculate the second moment:

$$\mathbb{E}(X^2) = \sum_{k=0}^{5} k^2 p_X(k)$$
$$= (25\cdot 2 + 16\cdot 4 + 9\cdot 6 + 4\cdot 8 + 1\cdot 10)/36 = \frac{35}{6}$$

Then, we can compute the variance:

$$\text{var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = 35/6 - (35/18)^2 = \frac{665}{324}$$

Example

You enter a game show where you can choose to answer the "easy" or "difficult" question first.

- The easy one: You win \$100 if answered correctly. You estimate to have probability 0.8 being correct.
- The difficult one: You win \$200 if answered correctly. You estimate to have probability 0.5 being correct.
- If you answer the first question correctly, you get a chance to attempt the second one. If not, you don't get to try the second question.

Which one should you try first, in order to maximize your expected return?

Example

Let X be the total prize money. Easy question first:

$$p_X(0) = 0.2 \qquad p_X(100) = 0.8\cdot 0.5 \qquad p_X(300) = 0.8\cdot 0.5$$
$$\mathbb{E}(X) = 0\cdot 0.2 + 100\cdot 0.8\cdot 0.5 + 300\cdot 0.8\cdot 0.5 = 160$$

Difficult question first:

$$p_X(0) = 0.5 \qquad p_X(200) = 0.5\cdot 0.2 \qquad p_X(300) = 0.5\cdot 0.8$$
$$\mathbb{E}(X) = 0\cdot 0.5 + 200\cdot 0.5\cdot 0.2 + 300\cdot 0.5\cdot 0.8 = 140$$

So easy one first in this case.

Example

Generalize to an optimal sequencing problem: Q1 with success p_1 and award v_1; Q2 with success p_2 and award v_2.

Q1 first: $\mathbb{E}(X) = v_1 p_1 (1 - p_2) + (v_1 + v_2) p_1 p_2 = p_1 v_1 + p_1 p_2 v_2$
Q2 first: $\mathbb{E}(X) = v_2 p_2 (1 - p_1) + (v_1 + v_2) p_2 p_1 = p_2 v_2 + p_1 p_2 v_1$

Go for Q1 if and only if:

$$p_1 v_1 + p_1 p_2 v_2 \geqslant p_2 v_2 + p_1 p_2 v_1$$
$$\frac{p_1 v_1}{1 - p_1} \geqslant \frac{p_2 v_2}{1 - p_2}$$

We can say $\dfrac{pv}{1-p}$ is our metric of which to try first.

1. Probability mass function
2. Expectation, mean, variance
3. Common discrete random variables
4. Joint PMF, conditioning, independence

Discrete random variable

It is useful to know the mean and variance of some common distributions. A summary:

	mean: $\mathbb{E}(X)$	variance: $\mathrm{var}(X)$
uniform: $X \sim U(a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)(b-a+2)}{12}$
Bernoulli: $X \sim \mathrm{Ber}(p)$	p	$p(1-p)$
Binomial: $X \sim \mathrm{Bin}(n, p)$	np	$np(1-p)$
Geometric: $X \sim \mathrm{Geo}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson: $X \sim \mathrm{Poisson}(\lambda)$	λ	λ

(1) Discrete uniform random variable

Discrete uniform random variable over a range $[a, b]$ has PMF

$$p_X(k) = \begin{cases} \dfrac{1}{b-a+1} & k = a, a+1, \ldots, b \\ 0 & \text{otherwise.} \end{cases}$$

Let $c = b - a$. Then,

$$\mathbb{E}(X) = \sum_x x\, p_X(x) = [(a) + (a+1) + \ldots + (a+c)] \cdot \frac{1}{c+1}$$
$$= \left[(c+1)a + \frac{c(c+1)}{2}\right] \cdot \frac{1}{c+1} = a + \frac{c}{2} = \frac{a+b}{2}$$

(1) Discrete uniform random variable

For variance, a "trick" is to consider a special case over $[1, n]$.

$$\mathbb{E}(X^2) = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{1}{6}(n+1)(2n+1)$$

$$\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2 = \frac{n^2-1}{12}$$

We now substitute $n = b - a + 1$, and note that shifting values does not change variance,

$$\text{var}(X) = \frac{(b-a+1)^2 - 1}{12} = \frac{(b-a)(b-a+2)}{12}$$

(2) Bernoulli random variable

Bernoulli random variable (p) has PMF

$$\boxed{p_X(k) = \begin{cases} p & k = 1 \\ 1-p & k = 0 \end{cases}}$$

$$\mathbb{E}(X) = \sum_x x\, p_X(x) = 1 \cdot p + 0 \cdot (1-p) = p$$

$$\mathbb{E}(X^2) = \sum_x x^2\, p_X(x) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$

$$\text{var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = p - p^2 = p(1-p)$$

$$\sigma_X = \sqrt{p(1-p)}$$

(3) Binomial random variable

Binomial random variable (n, p) has PMF

$$\boxed{p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n}$$

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} np \frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np
\end{aligned}$$

(3) Binomial random variable

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} k \frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k'=0}^{n-1} (k'+1) \binom{n-1}{k'} p^{k'} (1-p)^{(n-1)-(k')} \quad \text{where } k' = k-1 \\
&= np \left(\sum_{k'=0}^{n-1} (k') \binom{n-1}{k'} p^{k'} (1-p)^{(n-1)-(k')} \right) + np \\
&= np \cdot (n-1)p + np = np(1-p+np)
\end{aligned}$$

$$\text{var}(X) = np(1-p+np) - (np)^2 = np(1-p)$$

(4) Geometric random variable

Geometric random variable (p) has PMF

$$\boxed{p_X(k) = (1-p)^{k-1}p \qquad k = 1, 2, \ldots}$$

Let $S = \sum_{k=1}^{\infty} k(1-p)^{k-1}$

$$\begin{aligned} S &= 1 + 2(1-p) + 3(1-p)^2 + \ldots \\ (1-p)S &= \qquad (1-p) + 2(1-p)^2 + \ldots \end{aligned}$$

$$pS = S - (1-p)S = 1 + (1-p) + (1-p)^2 + \ldots = \frac{1}{1-(1-p)} = \frac{1}{p}$$

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

(4) Geometric random variable

$$\begin{aligned} T &= 1 + 4(1-p) + 9(1-p)^2 + \ldots \\ (1-p)T &= \qquad (1-p) + 4(1-p)^2 + \ldots \end{aligned}$$

$$\begin{aligned} T' = pT &= 1 + 3(1-p) + 5(1-p)^2 + \ldots \\ (1-p)T' &= \qquad (1-p) + 3(1-p)^2 + \ldots \end{aligned}$$

$$pT' = 1 + 2(1-p) + 2(1-p)^2 + \ldots = 1 + \frac{2(1-p)}{p}$$

$$\mathbb{E}(X^2) = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = pT = T' = \frac{1}{p} + \frac{2(1-p)}{p^2}$$

$$\mathrm{var}(X) = \frac{1}{p} + \frac{2(1-p)}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

(5) Poisson random variable

Poisson random variable (λ) has PMF

$$\boxed{p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad k = 0, 1, \ldots}$$

$$\begin{aligned} \mathbb{E}(X) = \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} &= \sum_{k=1}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} && k = 0 \text{ term is zero} \\ &= \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} && \text{now let } m = k-1 \\ &= \lambda \sum_{m=0}^{\infty} \frac{e^{-\lambda}\lambda^m}{m!} = \lambda \sum_m p_X(m) = \lambda \end{aligned}$$

(5) Poisson random variable

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

$$\mathrm{var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \lambda(\lambda+1) - \lambda^2 = \lambda$$

Overall comments

Not too difficult to calculate the mean and variance... but given the simplicity of the results, perhaps there is room for some "tricks".

- Remember the mean as the "representative value"
- Remember the (square root of) variance as the "spread"

Example

Let $X \sim \text{Poisson}(\lambda)$. Find $\mathbb{E}(1/(X+1))$.

ANS: Remember $\mathbb{E}(g(X)) = \sum_x g(x)p_X(x)$. Therefore,

$$\mathbb{E}(1/(X+1)) = \sum_{k=0}^{\infty} \frac{1}{k+1}\frac{e^{-\lambda}\lambda^k}{k!}$$

But we are not done yet, because we can simplify the form on the right. The key is to manipulate it to have a form where we can make use of the fact that the sum of Poisson PMF is one.

$$\begin{aligned}\sum_{k=0}^{\infty} \frac{1}{k+1}\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k+1)!} &= \frac{1}{\lambda}\sum_{k'=1}^{\infty} \frac{e^{-\lambda}\lambda^{k'}}{(k')!} \qquad \text{with } k' = k+1 \\ &= \frac{1}{\lambda}\left(\sum_{k'=0}^{\infty} \frac{e^{-\lambda}\lambda^{k'}}{(k')!} - \frac{e^{-\lambda}\lambda^0}{0!}\right) \\ &= \frac{1}{\lambda}\left(1 - e^{-\lambda}\right)\end{aligned}$$

1 Probability mass function

2 Expectation, mean, variance

3 Common discrete random variables

4 Joint PMF, conditioning, independence

Definition

Consider two r.v. X and Y, with joint PMF

$$\boxed{p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)} \tag{20}$$

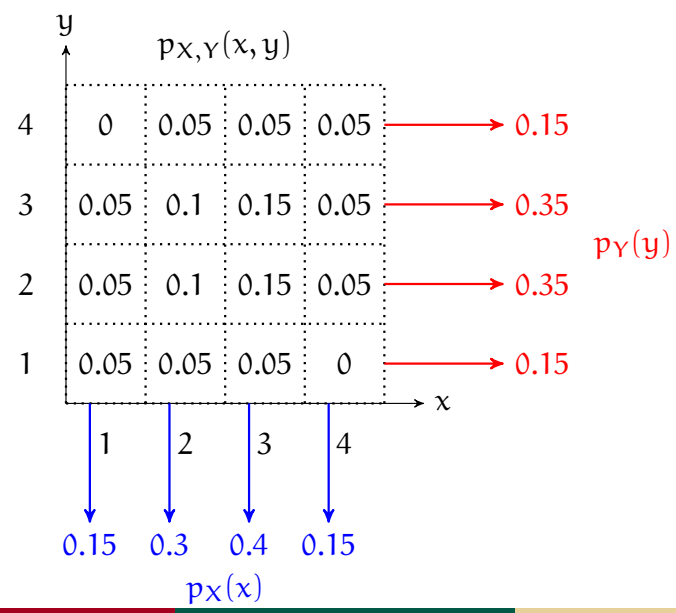
We can calculate the marginal PMFs of X and Y using

$$p_X(x) = \sum_y p_{X,Y}(x,y) \tag{21}$$

$$p_Y(y) = \sum_x p_{X,Y}(x,y) \tag{22}$$

and, of course, $\sum_x \sum_y p_{X,Y}(x,y) = 1$

Example



Extension

Some earlier formulas concerning single r.v. can be generalized: e.g., $Z = g(X, Y)$, then (cf. Eq. (8))

$$\boxed{p_Z(z) = \sum_{\{(x,y)\,|\,g(x,y)=z\}} p_{X,Y}(x,y)} \tag{23}$$

Expected value: (cf. Eq. (14))

$$\boxed{\mathbb{E}(g(X,Y)) = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)} \tag{24}$$

Linear function: (cf. Eq. (17))

$$\boxed{\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c} \tag{25}$$

From Bernoulli to binomial (Part 1)

The above can be generalized to multiple r.v.

Example: For a binomial r.v. $X = X_1 + X_2 + \ldots + X_n$, each trial is an (independent) Bernoulli r.v. X_i where $\mathbb{E}(X_i) = p$.

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np$$

We haven't quite made use of independence. Once we do, we will have a simple way to derive the variance as well. But first, we'll cover conditioning.

Conditioning on an event

We can have conditional PMF of a r.v. X given event A

$$\boxed{p_{X\,|\,A}(x) = \mathbb{P}(X = x\,|\,A) = \frac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)}} \tag{26}$$

similarly, $\sum_x p_{X\,|\,A}(x) = 1$.

Interpretation: Count all cases of $X = x$ AND belonging to A, then normalize with $\mathbb{P}(A)$.

Example

For a fair die, $p_X(k) = \frac{1}{6}$ for $k = 1, 2, 3, 4, 5, 6$. Let A be the event that the roll is odd. What is $p_{X|A}(x)$?

Ans:

$$p_{X|A}(k) = \frac{\mathbb{P}(\{X = k\} \cap \text{odd})}{\mathbb{P}(\text{odd})} = \begin{cases} \frac{1/6}{1/2} = \frac{1}{3} & k = 1, 3, 5 \\ 0 & \text{otherwise.} \end{cases}$$

Example

A student takes a test repeatedly up to a maximum of n times, each with probability p of passing, independent of number of attempts. What is the PMF of the number of attempts, given the student passes the test?

Ans: Let X be the r.v. if there is no limit of the number of attempts, so it is a *geometric r.v.* where $p_X(k) = (1-p)^{k-1}p$. Also, let $A = \{X \leqslant n\}$ be the event that the student passes the test. Thus, $\mathbb{P}(A) = \sum_{m=1}^{n}(1-p)^{m-1}p$.

$$p_{X|A}(k) = \begin{cases} \dfrac{(1-p)^{k-1}p}{\sum_{m=1}^{n}(1-p)^{m-1}p} & k = 1, 2, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

Conditioning on a r.v.

If event A is of the form $\{Y = y\}$, then we can have conditional PMF given another random variable.

The conditional PMF of a r.v. X given knowledge of another r.v. Y

$$\boxed{p_{X|Y}(x|y) = \mathbb{P}(X = x | Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}} \tag{27}$$

similarly, $\sum_x p_{X|Y}(x|y) = 1$.

Example

$p_{X,Y}(x, y)$

y \ x	1	2	3	4
4	0	0.05	0.05	0.05
3	0.05	0.1	0.15	0.05
2	0.05	0.1	0.15	0.05
1	0.05	0.05	0.05	0

$p_{X|Y}(x|4) = \{0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$

$p_{Y|X}(y|2) = \{\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\}$

From conditional to joint

We can "build up" the joint PMF if we have the conditionals, since

$$p_{X,Y}(x,y) = p_Y(y)\, p_{X|Y}(x|y)$$
$$p_{X,Y}(x,y) = p_X(x)\, p_{Y|X}(y|x)$$

We can also compute the marginal PMFs:

$$p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y p_Y(y)\, p_{X|Y}(x|y)$$
$$p_Y(y) = \sum_x p_{X,Y}(x,y) = \sum_x p_X(x)\, p_{Y|X}(y|x)$$

Example

You want to send messages over a computer network:

X : travel time of a given message

Y : length of the given message

We know the PMF of travel time of a message given its length, and the PMF of the length:

$$p_Y(y) = \begin{cases} \frac{5}{6} & y = 10^2 \\ \frac{1}{6} & y = 10^4 \end{cases}$$

$$p_{X|Y}(x|10^2) = \begin{cases} \frac{1}{2} & x = 10^{-2} \\ \frac{1}{3} & x = 10^{-1} \\ \frac{1}{6} & x = 1 \end{cases} \qquad p_{X|Y}(x|10^4) = \begin{cases} \frac{1}{2} & x = 1 \\ \frac{1}{3} & x = 10 \\ \frac{1}{6} & x = 100 \end{cases}$$

Find the PMF of the (unconditional) travel time a message.

Example

Ans:

$$p_X(x) = \sum_y p_Y(y)\, p_{X|Y}(x|y)$$
$$= \begin{cases} \frac{5}{6}\cdot\frac{1}{2} & x = 10^{-2} \\ \frac{5}{6}\cdot\frac{1}{3} & x = 10^{-1} \\ \frac{5}{6}\cdot\frac{1}{6} + \frac{1}{6}\cdot\frac{1}{2} & x = 1 \\ \frac{1}{6}\cdot\frac{1}{3} & x = 10 \\ \frac{1}{6}\cdot\frac{1}{6} & x = 10^2 \end{cases}$$

Conditional expectation

Let X and Y be two r.v., and A an event. We can compute the conditional expectations:

$$\mathbb{E}(X|A) = \sum_x x p_{X|A}(x) \tag{28}$$
$$\mathbb{E}(g(X)|A) = \sum_x g(x) p_{X|A}(x) \tag{29}$$
$$\mathbb{E}(X|Y) = \sum_x x p_{X|Y}(x|y) \tag{30}$$
$$\mathbb{E}(g(X)|Y) = \sum_x g(x) p_{X|Y}(x|y) \tag{31}$$

Conditional expectation

Assume $A_1, \ldots, A_n$ are disjoint events partitioning the sample space, with $\mathbb{P}(A_i) > 0$ for all i.

$$\boxed{\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(A_i)\,\mathbb{E}(X \mid A_i)} \tag{32}$$

unconditional average = averaging the conditional averages

Proof: Use $p_X(x) = \sum_{i=1}^{n} \mathbb{P}(A_i)\, p_{X \mid A_i}(x)$

$$\mathbb{E}(X) = \sum_x x p_X(x) = \sum_x x \sum_{i=1}^{n} \mathbb{P}(A_i)\, p_{X \mid A_i}(x) = \sum_{i=1}^{n} \mathbb{P}(A_i) \sum_x x p_{X \mid A_i}(x)$$
$$= \sum_{i=1}^{n} \mathbb{P}(A_i)\,\mathbb{E}(X \mid A_i)$$

Conditional expectation

Similarly, for any event B,

$$\boxed{\mathbb{E}(X \mid B) = \sum_{i=1}^{n} \mathbb{P}(A_i \mid B)\,\mathbb{E}(X \mid A_i \cap B)} \tag{33}$$

Also when $A_i = \{Y = y\}$,

$$\boxed{\mathbb{E}(X) = \sum_y p_Y(y)\,\mathbb{E}(X \mid Y = y)} \tag{34}$$

Example

Message can be transmitted to one of three cities:

- New York, with probability 0.5, mean time 0.05 second
- San Francisco, with probability 0.2, mean time 0.3 second
- Chicago, with probability 0.3, mean time 0.1 second

Find expected transmission time.

Ans: $\mathbb{E}(X) = 0.5 \cdot 0.05 + 0.2 \cdot 0.3 + 0.3 \cdot 0.1 = 0.115$ second.

Example

Find mean and variance of geometric distribution $p_X(k) = (1-p)^{k-1}p$.

Ans: Define $\begin{cases} A_1 = \{X = 1\} = \{\text{first trial is successful}\} \\ A_2 = \{X > 1\} = \{\text{first trial is unsuccessful}\} \end{cases}$

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{P}(A_1)\,\mathbb{E}(X \mid A_1) + \mathbb{P}(A_2)\,\mathbb{E}(X \mid A_2) \\ &= p \cdot 1 + (1-p) \cdot (\mathbb{E}(X) + 1) \end{aligned}$$

This gives $\mathbb{E}(X) = \frac{1}{p}$.

$$\begin{aligned} \mathbb{E}(X^2) &= \mathbb{P}(A_1)\,\mathbb{E}(X^2 \mid A_1) + \mathbb{P}(A_2)\,\mathbb{E}(X^2 \mid A_2) \\ &= p \cdot 1 + (1-p) \cdot \mathbb{E}\big((X+1)^2\big) \\ &= p + (1-p)(1 + 2\mathbb{E}(X) + \mathbb{E}(X^2)) \end{aligned}$$

This gives $\mathbb{E}(X^2) = \frac{2}{p^2} - \frac{1}{p}$, and therefore $\mathrm{var}(X) = \frac{1-p}{p^2}$.

Independence

Events $\{X = x\}$ and A are independent, i.e, r.v. X is independent of event A,

$$\boxed{\mathbb{P}(\{X = x\} \cap A) = \mathbb{P}(X = x)\,\mathbb{P}(A) = p_X(x)\,\mathbb{P}(A)} \tag{35}$$

Since $p_{X\,|\,A}(x) = \dfrac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)}$ by definition, independence means

$$\boxed{p_{X\,|\,A}(x) = p_X(x)} \tag{36}$$

Independence

Events $\{X = x\}$ and $\{Y = y\}$ are independent, i.e, r.v. X is independent of r.v. Y, if they have the relationship

$$\boxed{p_{X,Y}(x, y) = p_X(x)\,p_Y(y)} \tag{37}$$

Since $p_{X,Y}(x, y) = p_{X\,|\,Y}(x\,|\,y)\,p_Y(y)$, for all y with $p_Y(y) > 0$, we can conclude that

$$\boxed{p_{X\,|\,Y}(x\,|\,y) = p_X(x)} \tag{38}$$

Conditional independence

We can extend the concept to cases "conditional" on an event.

Two r.v. X and Y may have conditional independence given event A, if they satisfy

$$\mathbb{P}(X = x, Y = y\,|\,A) = \mathbb{P}(X = x\,|\,A)\,\mathbb{P}(Y = y\,|\,A)$$

which, using the notation of PMF,

$$\boxed{p_{X,Y\,|\,A}(x, y) = p_{X\,|\,A}(x)\,p_{Y\,|\,A}(y)} \tag{39}$$

In the same manner as the derivation above, the relationship is equivalent to

$$\boxed{p_{X\,|\,Y,A}(x\,|\,y) = p_{X\,|\,A}(x)} \tag{40}$$

Example: Conditional independence does not imply independence

X and Y are not independent.

Conditional on the event $A = \{X \leqslant 2 \cap Y \geqslant 3\}$, they are independent.

$p_{X,Y}(x, y)$

$y \backslash x$				
4	0.05	0.1	0.1	0
3	0.1	0.2	0.05	0.1
2	0	0.05	0.15	0.05
1	0	0.05	0	0

$$p_{X\,|\,A}(x) = \begin{cases} \frac{1}{3} & x = 1 \\ \frac{2}{3} & x = 2 \end{cases}$$

$$p_{Y\,|\,A}(y) = \begin{cases} \frac{1}{3} & y = 4 \\ \frac{2}{3} & y = 3 \end{cases}$$

$$p_{X,Y\,|\,A}(x, y) = \begin{cases} \frac{1}{9} & x = 1, y = 4 \\ \frac{2}{9} & x = 1, y = 3 \\ \frac{2}{9} & x = 2, y = 4 \\ \frac{4}{9} & x = 2, y = 3 \end{cases}$$

Example: Independence does not imply conditional independence

Two independent fair coin tosses, X and Y. In each case the payoff is \$1 for Head. For event $A = \{\text{total payoff} = \$1\}$, are they conditionally independent?

ANS:

$$p_{X|A}(x) = \begin{cases} \frac{1}{2} & x = 0 \\ \frac{1}{2} & x = 1 \end{cases}$$

$$p_{Y|A}(x) = \begin{cases} \frac{1}{2} & y = 0 \\ \frac{1}{2} & y = 1 \end{cases}$$

$$p_{X,Y|A}(x) = \begin{cases} \frac{1}{2} & x = 0, y = 1 \\ \frac{1}{2} & x = 1, y = 0 \end{cases}$$

So $p_{X,Y|A}(x,y) \neq p_{X|A}(x)\, p_{Y|A}(y)$.

Independence and expectation

If X and Y are independent, then

$$\boxed{\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}(X)\,\mathbb{E}(Y) && (41) \\ \mathbb{E}(g(X)h(Y)) &= \mathbb{E}(g(X))\,\mathbb{E}(h(Y)) && (42) \end{aligned}}$$

Proof (first case for simplicity):

$$\begin{aligned} \mathbb{E}(XY) &= \sum_x \sum_y xy\, p_{X,Y}(x,y) \\ &= \sum_x \sum_y xy\, p_X(x)\, p_Y(y) \qquad \text{(independence)} \\ &= \sum_x x p_X(x) \sum_y y p_Y(y) = \mathbb{E}(X)\,\mathbb{E}(Y) \end{aligned}$$

Independence and variance

If X and Y are independent, then

$$\boxed{\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)} \qquad (43)$$

Proof: Let $\tilde{X} = X - \mathbb{E}(X)$ and $\tilde{Y} = Y - \mathbb{E}(Y)$. Then, $\mathbb{E}\left(\tilde{X} + \tilde{Y}\right) = 0$ and $\mathbb{E}\left(\tilde{X}\tilde{Y}\right) = \mathbb{E}\left(\tilde{X}\right)\mathbb{E}\left(\tilde{Y}\right) = 0$.

$$\begin{aligned} \mathrm{var}(X+Y) = \mathrm{var}\left(\tilde{X}+\tilde{Y}\right) &= \mathbb{E}\left((\tilde{X}+\tilde{Y})^2\right) - \left(\mathbb{E}\left(\tilde{X}+\tilde{Y}\right)\right)^2 \\ &= \mathbb{E}\left(\tilde{X}^2 + 2\tilde{X}\tilde{Y} + \tilde{Y}^2\right) \\ &= \mathbb{E}\left(\tilde{X}^2\right) + 2\mathbb{E}\left(\tilde{X}\tilde{Y}\right) + \mathbb{E}\left(\tilde{Y}^2\right) \\ &= \mathbb{E}\left(\tilde{X}^2\right) + \mathbb{E}\left(\tilde{Y}^2\right) \\ &= \mathrm{var}\left(\tilde{X}\right) + \mathrm{var}\left(\tilde{Y}\right) = \mathrm{var}(X) + \mathrm{var}(Y) \end{aligned}$$

From Bernoulli to binomial (Part 2)

Example: For a binomial r.v. $X = X_1 + X_2 + \ldots + X_n$, each trial is an (independent) Bernoulli r.v. X_i where $\mathbb{E}(X_i) = p$. We already showed that $\mathbb{E}(X) = np$.

Similarly,

$$\mathrm{var}(X) = \mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{var}(X_i) = np(1-p)$$

Example

Let $X \sim \text{Ber}(p)$ and $Y \sim \text{Ber}(q)$, and that X and Y are independent. Find $\mathbb{E}\big((X+Y)^3\big)$.

ANS: For a Bernoulli random variable, we have derived that $\mathbb{E}(X) = \mathbb{E}\big(X^2\big) = p$. Similarly, $\mathbb{E}\big(X^3\big) = 1^3 \cdot p + 0^3 \cdot (1-p) = p$. We can also derive that

$$\begin{aligned}
\mathbb{E}\big((X+Y)^3\big) &= \mathbb{E}\big(X^3 + 3X^2Y + 3XY^2 + Y^3\big) \\
&= \mathbb{E}\big(X^3\big) + 3\mathbb{E}\big(X^2\big)\,\mathbb{E}(Y) + 3\mathbb{E}(X)\,\mathbb{E}\big(Y^2\big) + \mathbb{E}\big(Y^3\big) \\
&= p + 3pq + 3pq + q \\
&= p + q + 6pq
\end{aligned}$$

Conclusions

We developed the basic toolset for discrete random variables, and looked at some common distributions.

Next: Continuous random variables