Discrete Random Variables

Chia-Ping Chen

Professor
Department of Computer Science and Engineering
National Sun Yat-sen University

Probability

Outline

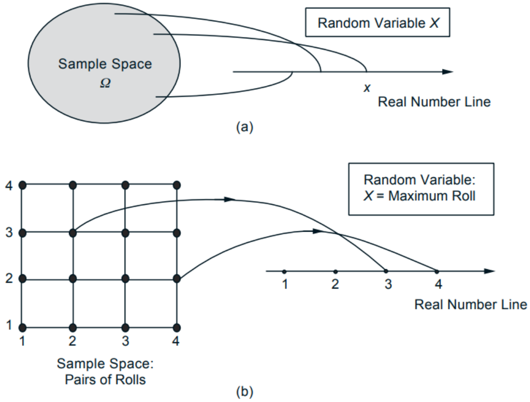
- Random variables
- Discrete random variables
- Expectation and variance
- Multiple random variables
- Total expectation theorem
- Independence

Random Variables

Definition (random variable)

Let $(\Omega, \mathcal{F}, P)$ be a probability model. A random variable (RV), say X, is a function from domain Ω to the real number set $\mathbb{R}$

$$X(\omega) = x, \quad \text{where } \omega \in \Omega,\ x \in \mathbb{R}$$



Random variable as enumeration or summarization

- Enumerating the possible outcomes of a random experiment
- Summarizing aspects of the outcomes in sample space

A sample space with 2 elements can be enumerated by a random variable with 2 values, e.g. 0 and 1.

- head | tail
- men | women
- win | lose
- pass | fail
- yes | no
- busy | idle
- spin up | down
- real | fake

Distribution

Let $(\Omega, \mathcal{F}, P)$ be a probability model and X be a random variable defined on Ω.

- The **distribution** of X specifies the probabilities of events associated with X
- The probability law over Ω is converted to distribution of X
- Distribution is specified by probabilistic function
  - probability mass function (PMF)
  - probability density function (PDF)
  - cumulative distribution function (CDF)

Discrete Random Variables

Definition (discrete random variables)

Let $(\Omega, \mathcal{F}, P)$ be a probability model.

- Let X be a random variable defined on Ω with range $\mathcal{X}$
- X is a discrete random variable (DRV) if $\mathcal{X}$ is countable

Definition (probability mass function)

Let $(\Omega, \mathcal{F}, P)$ be a probability model and X be a DRV defined on Ω with range $\mathcal{X}$.

- $X = x$ is an event for any $x \in \mathbb{R}$
- The probability mass function (PMF) of X is

$$p_X(x) = P(X = x) = \begin{cases} P(\{\omega \in \Omega \mid X(\omega) = x\}), & x \in \mathcal{X} \\ 0, & \text{otherwise} \end{cases}$$

- PMF is non-negative

$$P(X = x) \geq 0 \ \Rightarrow \ p_X(x) \geq 0$$

- PMF is normalized

$$\bigcup_{x \in \mathcal{X}} (X = x) = \Omega \ \Rightarrow \ \sum_{x \in \mathcal{X}} P(X = x) = P(\Omega) = 1$$

$$\Rightarrow \ \sum_{x \in \mathcal{X}} p_X(x) = 1$$

PMF from probability model

Let $\mathcal{M} = (\Omega, \mathcal{F}, P)$ be a probability model and X be a DRV defined on Ω. Both $\mathcal{X}$ and p_X can be derived.

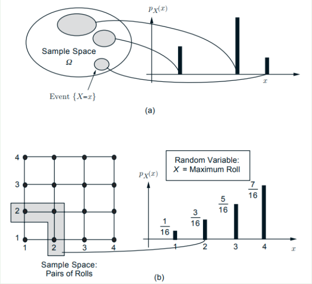
Example

Let X be the maximum of two rolls of a fair 4-face dice.

$$\Omega = \{1,2,3,4\} \times \{1,2,3,4\}$$

$$\mathcal{X} = \{1,2,3,4\}$$

$$p_X(x) = P(X = x) = \begin{cases} \frac{1}{16}, & x = 1 \\ \frac{3}{16}, & x = 2 \\ \frac{5}{16}, & x = 3 \\ \frac{7}{16}, & x = 4 \end{cases}$$



Basic random variables

- uniform
- Bernoulli
- binomial
- geometric
- Poisson

Uniform

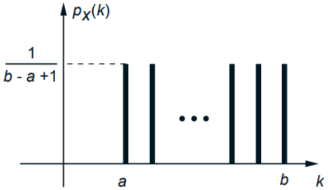
A uniform DRV has a constant PMF over its range.

- The uniform DRV taking an integer value from $a \in \mathbb{Z}$ to $b \in \mathbb{Z}$ has PMF

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

- This is denoted by

$$X \sim \textbf{Uni}[a, b]$$



Bernoulli

- A Bernoulli random variable X has PMF

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

  Note $p_X(\cdot)$ has a parameter $0 \leq p \leq 1$.

- This is denoted by

$$X \sim \textbf{Ber}(p)$$

- Note the PMF can be written as

$$p_X(k) = \begin{cases} p^k (1-p)^{1-k}, & k = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

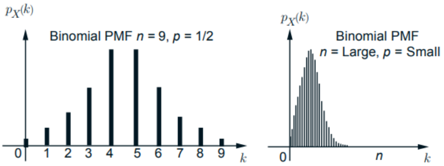
Binomial

- A binomial random variable has PMF

$$p_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0, 1, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

  Note $p_X(\cdot)$ has parameters $n \in \mathbb{N}$ and $0 \leq p \leq 1$.
- This is denoted by

$$X \sim \textbf{Bin}(n, p)$$



Geometric

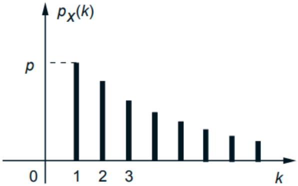
- A geometric random variable has PMF

$$p_X(k) = \begin{cases} (1-p)^{k-1}p, & k = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

  Note $p_X(\cdot)$ has parameter $0 \leq p \leq 1$.

- This is denoted by

$$X \sim \textbf{Geo}(p)$$



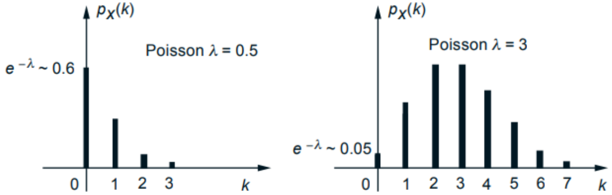
Poisson

- A Poisson random variable has PMF

$$p_X(k) = \begin{cases} e^{-\lambda}\frac{\lambda^k}{k!}, & k = 0, 1, \ldots \\ 0, & \text{otherwise} \end{cases}$$

  Note $p_X(\cdot)$ has parameter $\lambda > 0$.
- This is denoted by

$$X \sim \textbf{Poi}(\lambda)$$



Binomial and Poisson

Let Y be **Bin**(n, p) and $Z \sim$ **Poi**(np). For $n \gg 1$ and $p \approx 0$, the PMFs satisfy

$$p_Y(k) \approx p_Z(k)$$

for small k.

For $k \ll n$, we have $n - k \approx n$. Therefore

$$p_Y(k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{(np)^k}{k!} e^{-np} = p_Z(k)$$

The approximations used above are

$$\binom{n}{k} p^k \approx \frac{(np)^k}{k!}$$

$$(1-p)^{n-k} \approx (1-p)^n = \left((1-p)^{-\frac{1}{p}}\right)^{-np} \approx e^{-np}$$

Basic random variables and coin flips

Uni$[0,1]$	$\leftrightarrow$	# heads in a fair coin flip
Ber(p)	$\leftrightarrow$	# heads in a coin flip
Geo(p)	$\leftrightarrow$	# flips until the first head shows up
Bin(n,p)	$\leftrightarrow$	# heads in n coin flips
Poi(np)	$\leftrightarrow$	# heads in n coin flips with rare heads

Expectation

Definition (Expectation)

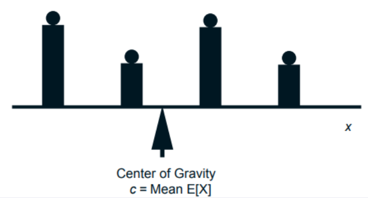
Let X be a DRV with range $\mathcal{X}$ and PMF p_X.

- The expectation (or expected value, mean) of X is defined by

$$\mathbf{E}[X] = \sum_{x \in \mathcal{X}} x\, p_X(x)$$

- $\mathbf{E}[X]$ is the center of gravity of the "probability mass"

$$\sum_{x \in \mathcal{X}} (x - \mathbf{E}[X])\, p_X(x) = \mathbf{E}[X] - \mathbf{E}[X] = 0$$



Example (2.2)

Two coins, each with probability $3/4$ for head, are tossed. Let X be the number of heads obtained. Decide p_X and $\mathbf{E}[X]$.

The range of X is $\mathcal{X} = \{0, 1, 2\}$. We have

$$p_X(0) = P(TT) = \left(\frac{1}{4}\right)^2$$
$$p_X(1) = P(HT) + P(TH) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{4}\right)\left(\frac{3}{4}\right)$$
$$p_X(2) = P(HH) = \left(\frac{3}{4}\right)^2$$

Hence

$$\mathbf{E}[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) = 0 + \frac{6}{16} + \frac{18}{16} = \frac{3}{2}$$

Example (2.4)

If the weather is good, with probability 0.6, Alice walks the 2 miles to school at a speed of 5 miles per hour. Otherwise, she rides her motorcycle at a speed of 30 miles per hour. What is the mean time for Alice to get to school?

Let V and T be speed and time. The ranges for V and T are $\mathcal{V} = \{5, 30\}$ and $\mathcal{T} = \left\{\frac{2}{5}, \frac{2}{30}\right\}$. We have

$$p_V(5) = p_T\left(\frac{2}{5}\right) = 0.6, \; p_V(30) = p_T\left(\frac{2}{30}\right) = 0.4$$

Hence

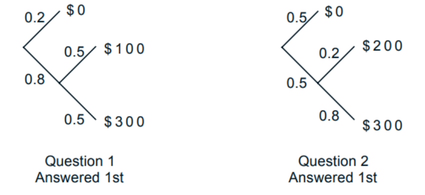
$$\mathbf{E}[T] = \sum_{t \in \mathcal{T}} t\, p_T(t) = \frac{2}{5} \cdot 0.6 + \frac{2}{30} \cdot 0.4 = \frac{4}{15}$$

Note that

$$\mathbf{E}\left[\frac{2}{V}\right] = \mathbf{E}[T] = \frac{4}{15} \neq \frac{2}{15} = \frac{2}{\mathbf{E}[V]}$$

Example (2.8)

There are questions Q1 and Q2 in a game. The probabilities of answering Q1 and Q2 correctly are 0.8 and 0.5. The rewards of answering Q1 and Q2 correctly are 100 and 200. If the question answered first is answered correctly, the remaining question can be answered. Which question should be answered first to maximize the expected reward?



Let R_1 and R_2 be the rewards when answering Q1 and Q2 first.

$$\mathbf{E}[R_1] = 0 \cdot 0.2 + 100 \cdot 0.4 + 300 \cdot 0.4 = 160, \ \mathbf{E}[R_2] = 140$$

Basic expectations

$U \sim \textsf{Uni}[a,b]$	$\mathbf{E}[U] = \frac{a+b}{2}$
$X \sim \textsf{Ber}(p)$	$\mathbf{E}[X] = p$
$Y \sim \textsf{Bin}(n,p)$	$\mathbf{E}[Y] = np$
$G \sim \textsf{Geo}(p)$	$\mathbf{E}[G] = \frac{1}{p}$
$Z \sim \textsf{Poi}(\lambda)$	$\mathbf{E}[Z] = \lambda$

Uniform expectation

Case $U \sim \textbf{Uni}[a, b]$.

$$\begin{aligned}
\mathbf{E}[U] &= \sum_{k=a}^{b} k\, p_U(k) \\
&= \left(\frac{1}{b-a+1}\right) \sum_{k=a}^{b} k \\
&= \left(\frac{1}{b-a+1}\right) \frac{(a+b)(b-a+1)}{2} \\
&= \frac{a+b}{2}
\end{aligned}$$

Bernoulli expectation

Case $X \sim \textbf{Ber}(p)$.

$$\begin{aligned} \mathbf{E}[X] &= 0 \cdot p_X(0) + 1 \cdot p_X(1) \\ &= 0 \cdot (1-p) + 1 \cdot p \\ &= p \end{aligned}$$

Binomial expectation

Case $Y \sim \textbf{Bin}(n, p)$.

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

Geometric expectation

Case $G \sim \textbf{Geo}(p)$.

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

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

$$\Rightarrow \sum_{k=1}^{\infty} (1-p)^{k} - \sum_{k=1}^{\infty} kp(1-p)^{k-1} = -1$$

$$\Rightarrow \sum_{k=1}^{\infty} kp(1-p)^{k-1} = \sum_{k=1}^{\infty} (1-p)^{k} + 1$$

$$\Rightarrow \mathbf{E}[G] = \frac{1-p}{1-(1-p)} + 1 = \frac{1}{p}$$

Poisson expectation

Case $Z \sim \textbf{Poi}(\lambda)$.

$$\begin{aligned}
\mathbf{E}[Z] &= \sum_{k=0}^{\infty} k \, p_Z(k) \\
&= \sum_{k=0}^{\infty} k \left(e^{-\lambda} \frac{\lambda^k}{k!} \right) \\
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
&= e^{-\lambda} \sum_{k'=0}^{\infty} \frac{\lambda^{k'+1}}{k'!} \\
&= \lambda e^{-\lambda} \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}$$

Function of a random variable

Let $(\Omega, \mathcal{F}, P)$ be a probability model, X be a DRV defined on Ω, and $Y = g(X)$. Then Y is a DRV.

Let $\mathcal{Y}$ be the range of Y.

- $\mathcal{Y}$ is determined by $\mathcal{X}$ and $g(\cdot)$
- $|\mathcal{Y}| \leq |\mathcal{X}|$, so $\mathcal{Y}$ (and Y) is discrete

To determine the PMF of Y, let $S_y = \{x \in \mathcal{X} \mid g(x) = y\}$ for $y \in \mathcal{Y}$. We have

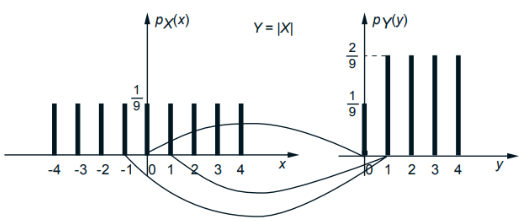
$$(Y = y) = \bigcup_{x \in S_y} (X = x)$$

$$\Rightarrow \; P(Y = y) = \sum_{x \in S_y} P(X = x)$$

$$\Rightarrow \; p_Y(y) = \sum_{x \in S_y} p_X(x)$$

EXAMPLE (2.1)

Let X be **Uni**$[-4, 4]$ and $Y = |X|$. Find p_Y.



- The range of Y is $\mathcal{Y} = \{0, 1, 2, 3, 4\}$.
- For $y \in \mathcal{Y}$, we determine $p_Y(y)$ through S_y.

Function expectation

Let X be a DRV with range $\mathcal{X}$ and PMF p_X. For any function $g(\cdot)$

$$\mathbf{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x)\, p_X(x)$$

Define $Y = g(X)$. We have

$$\begin{aligned}
\mathbf{E}[g(X)] &= \mathbf{E}[Y] \\
&= \sum_{y \in \mathcal{Y}} y\, p_Y(y) \\
&= \sum_{y \in \mathcal{Y}} y \sum_{x \in S_y} p_X(x) \\
&= \sum_{y \in \mathcal{Y}} \sum_{x \in S_y} y\, p_X(x) \\
&= \sum_{y \in \mathcal{Y}} \sum_{x \in S_y} g(x)\, p_X(x) \\
&= \sum_{x \in \mathcal{X}} g(x)\, p_X(x)
\end{aligned}$$

Linear function expectation

Let X be a DRV with range $\mathcal{X}$ and PMF p_X. The expectation of a linear function of X is

$$\mathbf{E}[aX+b] = a\,\mathbf{E}[X] + b$$

$$\begin{aligned}
\mathbf{E}[aX+b] &= \sum_{x\in\mathcal{X}} (ax+b)p_X(x) \\
&= \sum_{x\in\mathcal{X}} ax\,p_X(x) + \sum_{x\in\mathcal{X}} b\,p_X(x) \\
&= a\left(\sum_{x\in\mathcal{X}} x\,p_X(x)\right) + b\left(\sum_{x\in\mathcal{X}} p_X(x)\right) \\
&= a\,\mathbf{E}[X] + b
\end{aligned}$$

Definition (variance, standard deviation, moments)

Let X be a DRV with range $\mathcal{X}$ and PMF p_X.

- The variance of X, denoted by **var**(X), is the expectation of $(X - \mathbf{E}[X])^2$

$$\textbf{var}(X) = \mathbf{E}\left[(X - \mathbf{E}[X])^2\right]$$

- The standard deviation of X, denoted by σ_X, is the square root of the variance of X

$$\sigma_X = \sqrt{\textbf{var}(X)}$$

- The nth moment of X is the expectation of X^n

Linear function variance

Let X be a DRV. The variance of a linear function of X is

$$\textbf{var}(aX+b) = a^2\,\textbf{var}(X)$$

$$\begin{aligned}\textbf{var}(aX+b) &= \mathbf{E}\left[(aX+b-(a\mathbf{E}[X]+b))^2\right] \\ &= \mathbf{E}\left[(aX-a\mathbf{E}[X])^2\right] \\ &= a^2\,\mathbf{E}\left[(X-\mathbf{E}[X])^2\right] \\ &= a^2\,\textbf{var}(X)\end{aligned}$$

Variance formula

Let X be a DRV. We have

$$\textsf{var}(X) = \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X]$$

$$\begin{aligned}\textsf{var}(X) &= \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] \\ &= \mathbf{E}\left[X^2 - 2X\mathbf{E}[X] + \mathbf{E}^2[X]\right] \\ &= \mathbf{E}\left[X^2\right] - 2\mathbf{E}^2[X] + \mathbf{E}^2[X] \\ &= \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X]\end{aligned}$$

Example (2.3)

Let X be **Uni**$[-4, 4]$. Find the variance of X.

$$\textbf{var}(X) = \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X]$$
$$= \sum_{k=-4}^{4} k^2 p_X(k) - 0^2$$
$$= \frac{1}{9} \sum_{k=-4}^{4} k^2$$
$$= \frac{20}{3}$$

Basic variances

$$U \sim \textbf{Uni}[a,b] \qquad \textbf{var}(U) = \frac{(b-a)(b-a+2)}{12}$$

$$X \sim \textbf{Ber}(p) \qquad \textbf{var}(X) = p(1-p)$$

$$Y \sim \textbf{Bin}(n,p) \qquad \textbf{var}(Y) = np(1-p)$$

$$G \sim \textbf{Geo}(p) \qquad \textbf{var}(G) = \frac{1-p}{p^2}$$

$$Z \sim \textbf{Poi}(\lambda) \qquad \textbf{var}(Z) = \lambda$$

Uniform variance

Consider $U \sim \textbf{Uni}[a,b]$. Define $V = U - a$ so $V \sim \textbf{Uni}[0, b-a]$.

$$\begin{aligned}
\textbf{var}(U) &= \textbf{var}(V) \\
&= \mathbf{E}\left[V^2\right] - \mathbf{E}^2\left[V\right] \\
&= \left(\frac{1}{b-a+1}\right)\sum_{k=0}^{b-a} k^2 - \left(\frac{b-a}{2}\right)^2 \\
&= \left(\frac{1}{b-a+1}\right)\frac{(b-a)(b-a+1)(2(b-a)+1)}{6} - \left(\frac{b-a}{2}\right)^2 \\
&= \frac{(b-a)}{6}\left(2(b-a)+1-\frac{3}{2}(b-a)\right) \\
&= \frac{(b-a)(b-a+2)}{12}
\end{aligned}$$

Bernoulli variance

Consider $X \sim \textbf{Ber}(p)$. The second moment of X is

$$\begin{aligned}\mathbf{E}\left[X^2\right] &= 0^2 \cdot p_X(0) + 1^2 \cdot p_X(1) \\ &= 0^2 \cdot (1-p) + 1^2 \cdot p \\ &= p\end{aligned}$$

Thus

$$\begin{aligned}\textbf{var}(X) &= \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X] \\ &= p - p^2 \\ &= p(1-p)\end{aligned}$$

Geometric variance

Let G be **Geo**(p). We have shown earlier that

$$\mathbf{E}[G] = \sum_{k=1}^{\infty} k(1-p)^{k-1} p = \frac{1}{p}$$

Multiplying the sides by $(1-p)$ and taking the derivative with respect to p, we get

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

$$\mathbf{E}\left[G^2\right] = \sum_{k=1}^{\infty} k(1-p)^{k} + \frac{1}{p^2} = \frac{1-p}{p}\frac{1}{p} + \frac{1}{p^2} = \frac{2-p}{p^2}$$

Thus

$$\textsf{var}(G) = \mathbf{E}\left[G^2\right] - \mathbf{E}^2[G] = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Poisson variance

Consider $Z \sim \textbf{Poi}(\lambda)$. We have shown

$$\mathbf{E}[Z] = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \ \Rightarrow \ \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = \lambda e^{\lambda}$$

Taking the derivative with respect to λ, we have

$$\begin{aligned}
& \sum_{k=1}^{\infty} k^2 \frac{\lambda^{k-1}}{k!} = e^{\lambda} + \lambda e^{\lambda} \\
\Rightarrow \ & \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} = \lambda e^{\lambda} + \lambda^2 e^{\lambda} \\
\Rightarrow \ & \sum_{k=1}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \lambda + \lambda^2 \\
\Rightarrow \ & \mathbf{E}\left[Z^2\right] = \lambda + \lambda^2 \\
\Rightarrow \ & \textbf{var}(Z) = \mathbf{E}\left[Z^2\right] - \mathbf{E}^2[Z] = \lambda
\end{aligned}$$

Multiple Random Variables

Example (multiple random variables)

We may look at multiple aspects of the elements in a sample space.

1. date of birth of a person (Y, M, D)

$$Y = \text{year},\ M = \text{month},\ D = \text{day}$$

2. Hi-Life customers in one day (M, F)

$$M : \#\ \text{male customers},\ F : \#\ \text{female customers}$$

3. exit poll of a referendum $(X_1, \dots, X_n)$

$$X_i = \begin{cases} 1, & \text{pollee } i \text{ votes yes} \\ 0, & \text{otherwise} \end{cases}$$

Joint probability mass function

Let X and Y be DRVs with ranges $\mathcal{X}$ and $\mathcal{Y}$. The joint probability mass function (joint PMF) of X and Y specifies the probabilities over $\mathcal{X} \times \mathcal{Y}$. Specifically

$$p_{XY}(x, y) = \begin{cases} P(X = x \cap Y = y), & x \in \mathcal{X},\ y \in \mathcal{Y} \\ 0, & \text{otherwise} \end{cases}$$

- A joint PMF is non-negative

$$p_{XY}(x, y) \geq 0$$

- A joint PMF is normalized

$$\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) = 1$$

Example (Joint PMF)

1. date of birth

$$p_{YMD}(y,m,d) = P(Y=y \cap M=m \cap D=d)$$

2. Hi-Life customers

$$p_{MF}(m,f) = P(M=m \cap F=f)$$

Marginalization (sum rule)

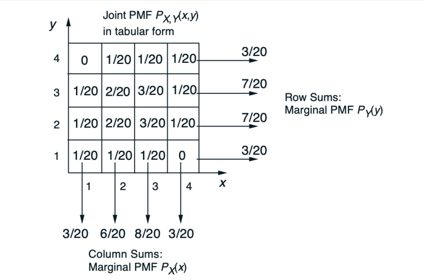
Let $(\Omega, \mathcal{F}, P)$ be a probability model. Let X and Y be DRVs defined on Ω, with ranges $\mathcal{X}$ and $\mathcal{Y}$ and joint PMF p_{XY}. Then

$$p_X(x) = \sum_{y \in \mathcal{Y}} p_{XY}(x, y)$$

$$\begin{aligned}
p_X(x) = P(X = x) &= P\left((X = x) \bigcap \Omega\right) \\
&= P\left((X = x) \bigcap \left(\bigcup_{y \in \mathcal{Y}} (Y = y)\right)\right) \\
&= P\left(\bigcup_{y \in \mathcal{Y}} ((X = x) \cap (Y = y))\right) \\
&= \sum_{y \in \mathcal{Y}} P(X = x \cap Y = y) \\
&= \sum_{y \in \mathcal{Y}} p_{XY}(x, y)
\end{aligned}$$

Example (2.9 Joint PMF)

Joint PMF of n DRVs can be represented by n-dimensional array.



By the sum rule, the marginal PMF p_Y consists of the row sums. Similarly, the marginal PMF p_X consists of the column sums.

Function of multiple random variables

A function of multiple DRVs is a DRV.

- Let X and Y be DRVs and $Z = g(X, Y)$. Then Z is a DRV.
- Range $\mathcal{Z}$ can be derived from $\mathcal{X}$ and $\mathcal{Y}$.
- To determine the PMF of Z, consider

$$S_z = \{(x, y) \in \mathcal{X} \times \mathcal{Y} \,|\, g(x, y) = z\}, \quad z \in \mathcal{Z}$$

We have

$$\begin{aligned} (Z = z) &= \bigcup_{(x,y)\in S_z} (X = x \cap Y = y) \\ \Rightarrow \; P(Z = z) &= \sum_{(x,y)\in S_z} P(X = x \cap Y = y) \\ \Rightarrow \; p_Z(z) &= \sum_{(x,y)\in S_z} p_{XY}(x, y) \end{aligned}$$

Function expectation

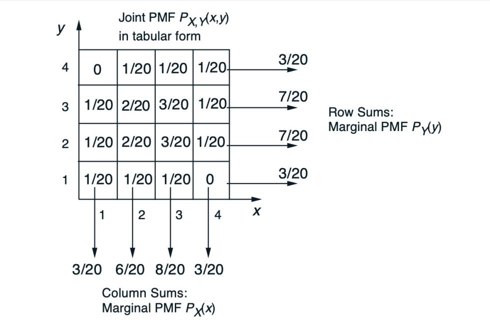
Let X and Y be DRVs and $g(\cdot,\cdot)$ be a function. Then

$$\mathbf{E}[g(X,Y)] = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} g(x,y)\, p_{XY}(x,y)$$

Consider $Z = g(X,Y)$. We have

$$\begin{aligned}
\mathbf{E}[g(X,Y)] = \mathbf{E}[Z] &= \sum_{z\in\mathcal{Z}} z\, p_Z(z) \\
&= \sum_{z\in\mathcal{Z}} z \sum_{(x,y)\in S_z} p_{XY}(x,y) \\
&= \sum_{z\in\mathcal{Z}} \sum_{(x,y)\in S_z} z\, p_{XY}(x,y) \\
&= \sum_{z\in\mathcal{Z}} \sum_{(x,y)\in S_z} g(x,y)\, p_{XY}(x,y) \\
&= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} g(x,y)\, p_{XY}(x,y)
\end{aligned}$$

Example (2.9 Expectation)



Find the expectation of $Z = X + 2Y$.

More than 2 random variables

Results with 2 random variables can be extended to more than two random variables.

- marginalization

$$p_{XY}(x,y) = \sum_{z\in\mathcal{Z}} p_{XYZ}(x,y,z)$$

$$p_X(x) = \sum_{y\in\mathcal{Y}} \sum_{z\in\mathcal{Z}} p_{XYZ}(x,y,z)$$

- expectation of a function of random variables

$$\mathbf{E}[g(X,Y,Z)] = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \sum_{z\in\mathcal{Z}} g(x,y,z) p_{XYZ}(x,y,z)$$

Linear function expectation

$$
\begin{aligned}
\mathbf{E}\left[\sum_{i=1}^{n} a_i X_i\right] &= \sum_{x_1\in\mathcal{X}_1}\cdots\sum_{x_n\in\mathcal{X}_n}\left(\sum_{i=1}^{n} a_i x_i\right) p_{X_1\ldots X_n}(x_1,\ldots,x_n) \\
&= \sum_{i=1}^{n} a_i\left[\sum_{x_1\in\mathcal{X}_1}\cdots\sum_{x_n\in\mathcal{X}_n} x_i\, p_{X_1\ldots X_n}(x_1,\ldots,x_n)\right] \\
&= \sum_{i=1}^{n} a_i\left[\sum_{x_i\in\mathcal{X}_i} x_i\left(\sum_{x_1\in\mathcal{X}_1}\cdots\sum_{x_n\in\mathcal{X}_n} p_{X_1\ldots X_n}(x_1,\ldots,x_n)\right)\right] \\
&= \sum_{i=1}^{n} a_i\left[\sum_{x_i\in\mathcal{X}_i} x_i\, p_{X_i}(x_i)\right] \\
&= \sum_{i=1}^{n} a_i\,\mathbf{E}[X_i]
\end{aligned}
$$

Binomial as a Bernoulli sum

Consider $B \sim \textbf{Bin}(n, p)$.

- B is the number of successes in n independent Bernoulli trials, where p is the probability of success for each trial
- Let X_i is the number of success for trial i (also known as an **indicator**). Then $X_i \sim \textbf{Ber}(p)$ and

$$B = X_1 + \ldots X_n$$

- The expectation of B is

$$\begin{aligned} \mathbf{E}[B] &= \mathbf{E}[X_1 + \cdots + X_n] \\ &= \mathbf{E}[X_1] + \cdots + \mathbf{E}[X_n] \\ &= np \end{aligned}$$

- The result agrees with the earlier direct derivation of $\mathbf{E}[B]$

Example (2.10 Independent Bernoulli sum)

Each student in a 300-student class has probability $1/3$ of getting a grade of A, independent of other students. What is the mean of X, the number of students getting A?

We have

$$X = X_1 + \cdots + X_{300}$$

where $X_i \sim \textbf{Ber}\left(\frac{1}{3}\right)$ is the indicator for student i getting A. Thus

$$\begin{aligned}
\mathbf{E}[X] &= \mathbf{E}[X_1 + \cdots + X_{300}] \\
&= \mathbf{E}[X_1] + \cdots + \mathbf{E}[X_{300}] \\
&= 300\left(\frac{1}{3}\right) \\
&= 100
\end{aligned}$$

Example (2.11 Dependent Bernoulli sum)

n persons put their hats in a box, and then each person randomly retrieves a hat. What is the expected value of H, the number of persons retrieving their own hats?

$H = X_1 + \cdots + X_n$, where X_i is the indicator for person i getting own hat. Thus

$$\begin{aligned}
\mathbf{E}[H] &= \sum_{i=1}^{n} \mathbf{E}[X_i] \\
&= \sum_{i=1}^{n} P(X_i = 1) \\
&= \sum_{i=1}^{n} P(\text{hat } i \text{ still in}) P(\text{picks hat } i \,|\, \text{hat } i \text{ still in}) \\
&= \sum_{i=1}^{n} \left(\frac{n-(i-1)}{n}\right)\left(\frac{1}{n-(i-1)}\right) = \sum_{i=1}^{n} \frac{1}{n} = 1
\end{aligned}$$

Conditional Probability of a Random Variable

PMF conditional on event

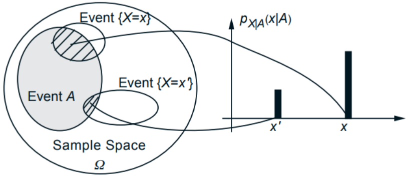
Let X be a DRV and A be an event with non-zero probability.

- Conditional on A, the probability of $(X = x)$ is

$$P(X = x \mid A) = \frac{P(X = x \cap A)}{P(A)}$$

- By definition, the conditional PMF of X given A, denoted by $p_{X|A}$, is

$$p_{X|A}(x) = P(X = x \mid A)$$



Example (2.12 Conditional PMF)

Let X be the roll of a fair six-face dice, and A be the event that X is even. What is $p_{X|A}$?

$$\begin{aligned} p_{X|A}(x) &= P(X = x \mid A) \\ &= \frac{P(X = x \cap A)}{P(A)} \\ &= \begin{cases} \frac{1}{3}, & x = 2, 4, 6 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

PMF conditional on random variable

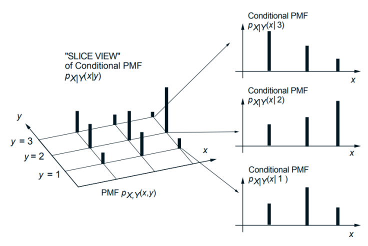
Let X and Y be DRVs. Note $(Y = y)$ is an event.

- The probability of $(X = x)$ given $(Y = y)$ is

$$P(X = x \,|\, Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}$$

- We define the conditional PMF of X given $(Y = y)$

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



PMF factorization

Let X, Y be DRVs. For $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, we have

$$P(X = x \cap Y = y) = P(X = x \,|\, Y = y) P(Y = y)$$

It follows that

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

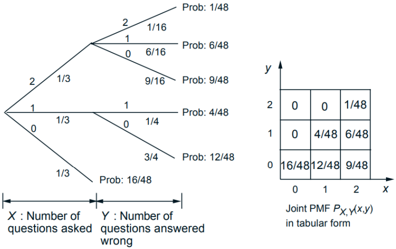
- Thus, joint PMF p_{XY} can be obtained via marginal PMF p_Y and conditional PMF $p_{X|Y}$ (or p_X and $p_{Y|X}$)
- Factorization is also known as the multiplication rule

Example (2.14 Factorization)

Professor May B. Right answers a question incorrectly with probability 1/4, independent of other questions. In a lecture, she is asked 0, 1, or 2 questions with equal probability 1/3. Let X be the number of questions she is asked, and Y be the number of questions she answers wrong in a lecture. Find $p_{XY}(x,y)$. What is the probability that she answers at least one question incorrectly?

We have

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y=0) \\ &= 1 - \sum_{x \in \mathcal{X}} p_{XY}(x,0) \\ &= 1 - \sum_{x \in \mathcal{X}} p_X(x) p_{Y|X}(0|x) \end{aligned}$$



Example (2.15 Factorization)

A transmitter sends messages over a computer network. Let Y be the length of a message, X be the transmission time, and

$$p_Y(y) = \begin{cases} \frac{5}{6}, & y = 10^2 \\ \frac{1}{6}, & y = 10^4 \\ 0, & \text{otherwise} \end{cases}, \quad p_{X|Y}(x|y) = \begin{cases} \frac{1}{2}, & x = 10^{-4}y \\ \frac{1}{3}, & x = 10^{-3}y \\ \frac{1}{6}, & x = 10^{-2}y \\ 0, & \text{otherwise} \end{cases}$$

Is X discrete? If so, what is the PMF of X?

From $\mathcal{Y} = \{10^2, 10^4\}$, the range of X is

$$\mathcal{X} = \{10^{-2}, 10^{-1}, 10^0\} \cup \{10^0, 10^1, 10^2\} = \{10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$$

So X is discrete. For the PMF of X, we have

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

Total Probability and Total Expectation

Total probability theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability model, $(A_1, \ldots, A_n)$ be a partition, and X be a DRV. Then

$$p_X(x) = \sum_{i=1}^{n} P(A_i)\, p_{X|A_i}(x)$$

$$\begin{aligned}
(X = x) &= (X = x \cap \Omega) \\
&= \left(X = x \cap \bigcup_{i=1}^{n} A_i \right) \\
&= \bigcup_{i=1}^{n} (X = x \cap A_i) \\
\Rightarrow \; P(X = x) &= \sum_{i=1}^{n} P(X = x \cap A_i) = \sum_{i=1}^{n} P(A_i) P(X = x | A_i) \\
\Rightarrow \; p_X(x) &= \sum_{i=1}^{n} P(A_i)\, p_{X|A_i}(x)
\end{aligned}$$

Total probability theorem

Let X and Y be DRVs. Then

$$p_X(x) = \sum_{y \in \mathcal{Y}} p_Y(y) p_{X|Y}(x|y)$$

The collection $\{(Y = y) \mid y \in \mathcal{Y}\}$ is a partition of sample space, so

$$P(X = x) = \sum_{y \in \mathcal{Y}} P(Y = y) P(X = x | Y = y)$$

Re-writing this with PMFs, we get

$$p_X(x) = \sum_{y \in \mathcal{Y}} p_Y(y) p_{X|Y}(x|y)$$

Note the consistency with marginalization

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

Definition (Conditional expectation)

Let X be a DRV. A **conditional expectation** of X is an expectation with respect to a conditional PMF of X.

- The expectation of X conditional on event A is

$$\mathbf{E}[X \mid A] = \sum_{x \in \mathcal{X}} x \, p_{X|A}(x)$$

- Let Y be a DRV. The expectation of X conditional on $(Y = y)$ is

$$\mathbf{E}[X \mid Y = y] = \sum_{x \in \mathcal{X}} x \, p_{X|Y}(x|y)$$

- Define the conditional expectation of X given Y by

$$\mathbf{E}[X|Y] = g(Y) \text{ where } g(y) = \mathbf{E}[X \mid Y = y]$$

Total expectation theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability model, $(A_1, \ldots, A_n)$ be a partition, and X be a DRV. Then

$$\mathbf{E}[X] = \sum_{i=1}^{n} P(A_i)\mathbf{E}[X|A_i]$$

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{x\in\mathcal{X}} x\, p_X(x) \\
&= \sum_{x\in\mathcal{X}} x \sum_{i=1}^{n} p_{X|A_i}(x) P(A_i) \\
&= \sum_{i=1}^{n} P(A_i) \sum_{x\in\mathcal{X}} x\, p_{X|A_i}(x) \\
&= \sum_{i=1}^{n} P(A_i)\, \mathbf{E}[X|A_i]
\end{aligned}$$

Total expectation theorem

Let X and Y be DRVs. The conditional expectation $\mathbf{E}[X|Y]$ is a random variable. The expectation of $\mathbf{E}[X|Y]$ is $\mathbf{E}[X]$.

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]]$$

$$\begin{aligned}
\mathbf{E}[\mathbf{E}[X|Y]] &= \sum_{y\in\mathcal{Y}} p_Y(y)\mathbf{E}[X|Y=y] \\
&= \sum_{y\in\mathcal{Y}} p_Y(y) \sum_{x\in\mathcal{X}} x p_{X|Y}(x|y) \\
&= \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} x p_Y(y) p_{X|Y}(x|y) \\
&= \sum_{x\in\mathcal{X}} x \sum_{y\in\mathcal{Y}} p_{XY}(x,y) \\
&= \sum_{x\in\mathcal{X}} x p_X(x) \\
&= \mathbf{E}[X]
\end{aligned}$$

Example (2.16 Total expectation)

Message transmitted by a computer in **Boston** through a data network is destined for **New York** with probability 0.5, for **Chicago** with probability 0.3, and for **San Francisco** with probability 0.2. The transmission time X is random. The mean transmission time is 0.05 seconds for a message destined for New York, 0.1 seconds for a message destined for Chicago, and 0.3 seconds for a message destined for San Francisco. What is $\mathbf{E}[X]$?

The partitioning events are

$$A_1 = \{\text{to NY}\},\ A_2 = \{\text{to Chicago}\},\ A_3 = \{\text{to SF}\}$$

We have

$$\begin{aligned}\mathbf{E}[X] &= \sum_{i=1}^{n} P(A_i)\mathbf{E}[X|A_i] \\ &= 0.5 \cdot 0.05 + 0.3 \cdot 0.1 + 0.2 \cdot 0.3\end{aligned}$$

Example (2.17 Total expectation)

A program written for a task works with probability p. What is the mean and variance of X, the number of tries until a program works?

Define $A = \{\text{first program works}\}$ and let $X' = X - 1$. Applying total expectation of X using A and A^c, we have

$$\mathbf{E}[X] = P(A)\mathbf{E}[X|A] + P(A^c)\mathbf{E}[X|A^c] = p \cdot 1 + (1-p)\mathbf{E}\left[1 + X'|A^c\right]$$

$$= p \cdot 1 + (1-p)(1 + \mathbf{E}\left[X'|A^c\right]) \xrightarrow{p_{X'|A^c} = p_X} \mathbf{E}[X] = \frac{1}{p}$$

$$\mathbf{E}\left[X^2\right] = P(A)\mathbf{E}\left[X^2|A\right] + P(A^c)\mathbf{E}\left[X^2|A^c\right]$$

$$= p \cdot 1^2 + (1-p)\mathbf{E}\left[(1+X')^2|A^c\right] \longrightarrow \mathbf{E}\left[X^2\right] = \frac{2-p}{p^2}$$

$$\Rightarrow \textbf{var}(X) = \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X] = \frac{1-p}{p^2}$$

Note $X \sim \textbf{Geo}(p)$ and the consistency with earlier results.

Poe divination

In poe divination, a.k.a. bwa bwei

- A divination seeker drops two little wooden pieces on the floor to get an answer represented by the positions of the pieces
- We assume the pieces are fair for simplicity
- We are interested in X, the random number of drops until back-to-back "divines" occur
- You can extend the analysis to back-to-back-to-back divines

$\mathbf{E}[X]$ via total expectation

Let the result of a drop of poe be D (for divine) or T. Use partitioning events $A_1 = \{\text{first drop is } T\}$, $A_2 = \{\text{first 2 drops are } DT\}$, and $A_3 = \{\text{first 2 drops are } DD\}$. By total expectation

$$\begin{aligned}
\mathbf{E}[X] &= P(A_1)\mathbf{E}[X|A_1] + P(A_2)\mathbf{E}[X|A_2] + P(A_3)\mathbf{E}[X|A_3] \\
&= \frac{1}{2}\left(\mathbf{E}\left[1 + X'|A_1\right]\right) + \frac{1}{4}\left(\mathbf{E}\left[2 + X''|A_2\right]\right) + \frac{1}{4}\mathbf{E}[2|A_3] \\
&= \frac{1}{2}\left(1 + \mathbf{E}\left[X'|A_1\right]\right) + \frac{1}{4}\left(2 + \mathbf{E}\left[X''|A_2\right]\right) + \frac{1}{4}\cdot 2 \\
&= \frac{1}{2}\left(1 + \mathbf{E}[X]\right) + \frac{1}{4}\left(2 + \mathbf{E}[X]\right) + \frac{1}{2}
\end{aligned}$$

since $p_{X''|A_2} = p_{X'|A_1} = p_X$. It follows that

$$\frac{1}{4}\mathbf{E}[X] = \frac{3}{2} \;\Rightarrow\; \mathbf{E}[X] = 6$$

$\mathbf{E}[X]$ via alternative total expectation

Use N, the number of drops until the first divine occurs, as conditioning random variable. We have

$$\begin{aligned}
\mathbf{E}[X] &= \mathbf{E}[\mathbf{E}[X|N]] \\
&= \mathbf{E}\left[\frac{1}{2}(N+1) + \frac{1}{2}(N+1+X')\right] \\
&= \mathbf{E}[N+1] + \frac{1}{2}\mathbf{E}[\mathbf{E}[X'|N]] \\
&= \mathbf{E}[N+1] + \frac{1}{2}\mathbf{E}[\mathbf{E}[X]] \text{ (since } p_{X'|N} = p_X) \\
&= \mathbf{E}[N] + 1 + \frac{1}{2}\mathbf{E}[X] \\
&= 2 + 1 + \frac{1}{2}\mathbf{E}[X]
\end{aligned}$$

Therefore

$$\mathbf{E}[X] = 6$$

PMF

Let T_n (resp. D_n) be the event of n drops ending in T (resp. D) still without a back-to-back divine. Note that T_n can be either D_{n-1} or T_{n-1} followed by a drop of T, while D_n must be T_{n-1} followed by a drop of D. That is

$$P(T_n) = P(T_{n-1})P(T) + P(D_{n-1})P(T) = \frac{1}{2}P(T_{n-1}) + \frac{1}{2}P(D_{n-1})$$

$$P(D_n) = P(T_{n-1})P(D) = \frac{1}{2}P(T_{n-1})$$

Event $(X = n)$ is D_{n-1} followed by D, so

$$P(X = n) = \frac{1}{2}P(D_{n-1})$$

For a small n, we can compute $P(X = n)$ recursively starting from $P(D_1) = P(T_1) = \frac{1}{2}$. A general formula of $P(X = n)$ can be worked out from the recursion relation (exercise).

Independence

Independence

Let X be DRV and A be an event. By definition, X is **independent** of A if $(X = x)$ is independent of A for every $x \in \mathcal{X}$.

The independence of X and A is denoted by

$$X \perp\!\!\!\perp A$$

For $X \perp\!\!\!\perp A$, we have

$$P(X = x \cap A) = P(X = x)P(A)$$

Dividing both sides by $P(A)$, we have

$$p_{X|A}(x) = p_X(x)$$

Example (2.19)

Flip a fair coin twice. Let X be the number of heads and A be the event that X is even. Then X and A are not independent.

Independence of random variables

Let X and Y be DRVs. By definition, X and Y are independent if

$$p_{XY}(x,y) = p_X(x)p_Y(y)$$

The independence of X and Y is denoted by

$$X \perp\!\!\!\perp Y$$

For $X \perp\!\!\!\perp Y$, we have

$$P(X = x \cap Y = y) = P(X = x)P(Y = y)$$

Dividing both sides by $P(Y = y)$, we have

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

Conditional independence

Let X, Y, Z be DRVs and A be an event.

- X and Y are conditionally independent given A if $(X = x)$ and $(Y = y)$ are independent given A for every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$
- X and Y are conditionally independent given Z if X and Y are conditionally independent given $(Z = z)$ for every $z \in \mathcal{Z}$

The conditional independence of X and Y given A is denoted by

$$X \perp\!\!\!\perp Y \,|\, A$$

The conditional independence of X and Y given Z is denoted by

$$X \perp\!\!\!\perp Y \,|\, Z$$

Simplification by independence

Suppose $X \perp\!\!\!\perp Y$. We have

$$\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\mathbf{E}[h(Y)]$$

$$\textbf{var}(X+Y) = \textbf{var}(X) + \textbf{var}(Y)$$

$$\begin{aligned}
\mathbf{E}[g(X)h(Y)] &= \sum_{x,y} p_{XY}(x,y)g(x)h(y) \\
&= \sum_{x,y} p_X(x)p_Y(y)g(x)h(y) \\
&= \sum_{x} p_X(x)g(x) \sum_{y} p_Y(y)h(y) \\
&= \mathbf{E}[g(X)]\mathbf{E}[h(Y)]
\end{aligned}$$

$$\begin{aligned}
\textbf{var}(X+Y) &= \mathbf{E}\left[(X+Y)^2\right] - \mathbf{E}^2[X+Y] \\
&= \mathbf{E}\left[X^2 + 2XY + Y^2\right] - (\mathbf{E}[X] + \mathbf{E}[Y])^2 \\
&= \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X] + \mathbf{E}\left[Y^2\right] - \mathbf{E}^2[Y] \\
&= \textbf{var}(X) + \textbf{var}(Y)
\end{aligned}$$

Example (2.20 Independent Bernoulli sum)

Consider $B \sim \textbf{Bin}(n, p)$. The variance of B can be derived as follows.

- B is the sum of n independent $\textbf{Ber}(p)$

$$B = X_1 + \cdots + X_n$$

- By the previous slide

$$\begin{aligned}\textbf{var}(B) &= \textbf{var}(X_1) + \cdots + \textbf{var}(X_n) \\ &= \sum_{i=1}^{n} p(1-p) \\ &= np(1-p)\end{aligned}$$

Note the consistency with earlier results.

Example (2.21 Independent Bernoulli average)

The approval rating of a politician can be estimated by asking voters randomly drawn from the voter population, called a poll. Let X_i indicate whether the ith asked voter approves the politician. Then the approval rating based on the poll is

$$R_n = \frac{X_1 + \cdots + X_n}{n}$$

Assume $X_1, \ldots, X_n$ are independent **Ber**(p), where p is the unknown approval rating. Find the mean and variance of R_n.

Example (2.22 Independent Bernoulli average)

Let A be an event. The probability of A can be estimated with the relative frequency of A in a simulation consisting of n independent runs of a random experiment.

- Let X_i indicates whether A occurs in run i
- The relative frequency of A in n runs is

$$F_n = \frac{X_1 + \cdots + X_n}{n}$$

- It follows that

$$\mathbf{E}[F_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[X_i] = P(A)$$

$$\textbf{var}(F_n) = \frac{1}{n^2}\sum_{i=1}^{n}\textbf{var}(X_i) = \frac{1}{n}[P(A)(1-P(A))] \xrightarrow{n\to\infty} 0$$

Probability mass function (PMF)

$$p_X(x) = P(X = x)$$

Expectation

$$\mathbf{E}[X] = \sum_x x\, p_X(x)$$

Basic DRVs

$$\textbf{Uni}[a, b],\ \textbf{Ber}(p),\ \textbf{Bin}(n, p)$$

$$\textbf{Geo}(p),\ \textbf{Poi}(\lambda)$$

Summary 2

Joint PMF

$$p_{XY}(x, y) = P(X = x \cap Y = y)$$

Marginalization

$$p_X(x) = \sum_y p_{XY}(x, y)$$

Conditional PMF

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

$$p_{X|A}(x) = P(X = x|A)$$

Summary 3

Total probability theorem

$$p_X(x) = \sum_i P(A_i) p_{X|A_i}(x)$$

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

Total expectation theorem

$$\mathbf{E}[X] = \sum_i P(A_i) \mathbf{E}[X|A_i]$$

$$\mathbf{E}[X] = \sum_y p_Y(y) \mathbf{E}[X|Y=y] = \mathbf{E}[\mathbf{E}[X|Y]]$$