# Winter 2024

# STA 347

By 面条

# Winter 2024 STA 347

February 11, 2024

## Contents

# 1 Preliminary

## 1.1 Probability Models

**Probability measure (三条公理)**

i) $\forall A \subset \Omega$, $0 \leq P(A) \leq 1$;

ii) $P(\emptyset) = 0, P(\Omega) = 1$;

iii) $\forall A_1, A_2, \ldots \subset \Omega$ such that $A_i \cap A_j = \emptyset \ \forall i \neq j$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

**注意：这里一定得保证是 disjoint union**

## 1.2 Conditional Probability

For any two events $A$ and $B$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B)\,P(B)$$

注意如果我们是 condition on event $B$ 发生，那么分母上对应的也是 event $B$ 的概率。

**Definition 1.1** Two events $A$ and $B$ are said to be *independent* if

$$P(A \cap B) = P(A)P(B)$$

## 1.3 Law of Total Probability

**Law of Total Probability**

Suppose $B_1, B_2, \ldots$ are disjoint events such that $\bigcup_{\forall i} B_i = \Omega$(sample space), we can consider $B_i$ as a **partition** of our sample space, then

$$P(A) = \sum_{\forall i} P(A \cap B_i) = \sum_{\forall i} P(A|B_i) P(B_i)$$

通常来说，当我们没法直接求出事件 $A$ 的概率，但是我们知道事件 $A$ given 某一个事件 $B_i$ 的概率，我们会考虑用 Law of total probability 进行计算。

## 1.4 Bayes' Formula

Bayes' Formula 应该算是求解 conditional probability 最常用的一个方法了。回到我们前面那张图，如果我现在知道了事件 $A$ 发生，我想知道它发生在哪一个 $B_i$ 的区域里？也就是说 $P(B_i|A)$?

**Bayes's Formula**

Suppose $B_1, B_2, \ldots$ forms a **partition** of sample space $\Omega$, and $A$ is any event with $P(A) \neq 0$, we have

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i) P(B_i)}{\sum_{\forall i} P(A|B_i) P(B_i)}$$

## 1.5 Limiting Events

当我们把一个实验重复进行多次，我们一般就会得到 a sequence of events。

**Example 1.1** Suppose a fair coin is flipped consecutively, we are interested in the following event

$$A = \{\text{“Head” never seen}\},$$

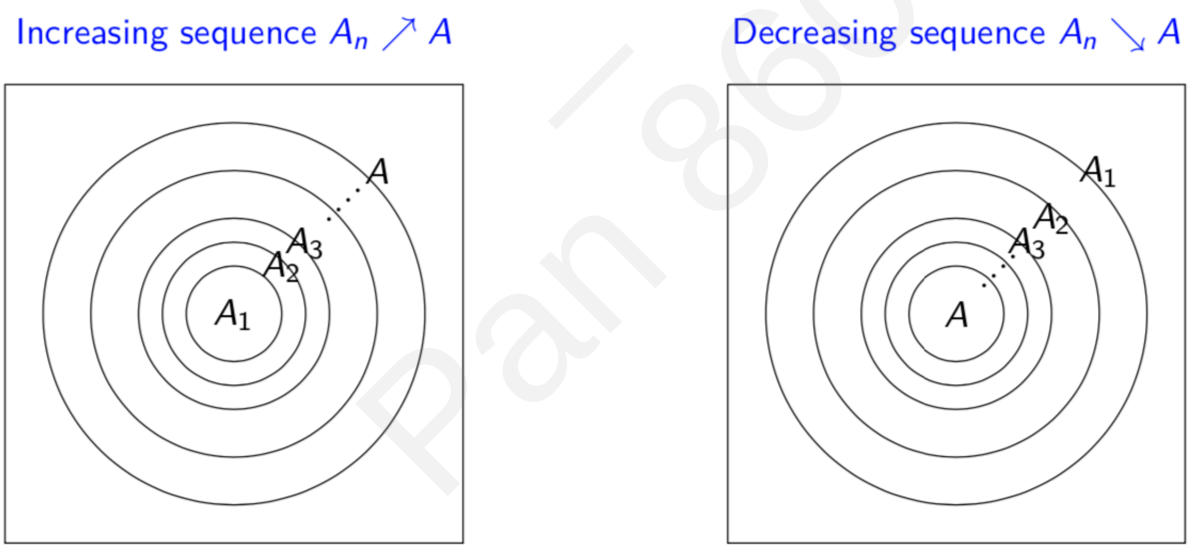
which is just the intersection, $A = \bigcap_{n\geq 1} A_n$ of the events

$$A_n = \{\text{“Heads” not seen in the first n tosses}\}.$$

上面这个例子里面，我们可以把 $A$ 想象成 $A_n$ 的 limit。不过想要严格定义 limit of events 比较复杂，所以我们先考虑 **monotone** sequence of events.

- Consider a sequence of events $A_1, A_2, \ldots \subset \Omega$, if $A_n \subset A_{n+1}$ for all $n \geq 1$, this sequence is **increasing**, denoted as $\{A_n\} \nearrow A = \bigcup_{n\geq 1} A_n$.
- Consider a sequence of events $A_1, A_2, \ldots \subset \Omega$, if $A_n \supseteq A_{n+1}$ for all $n \geq 1$, this sequence is **decreasing**, denoted as $\{A_n\} \searrow A = \bigcap_{n\geq 1} A_n$.

刚才上面举的那个例子可以发现，$\{A_n\}$ is decreasing.



**Continuity in Probability**

If $\{A_n\} \nearrow A$ or $\{A_n\} \searrow A$, then

$$\lim_{n\to\infty} P(A_n) = P(A).$$

# 2 Random Variables (RV)

## 2.1 Univariate rv

$$\text{RV}\begin{cases}\text{Discrete: Possible value in the form } \{x_1, x_2, \ldots\} \\ \text{Continuous: Possible value contains an interval, e.g. [2,5].}\end{cases}$$

**Cumulative distribution function (cdf)**

$$F(x) = P(X \leq x)$$

i) For all $x \leq y$, $F(x) \leq F(y)$ (**increasing** function)

ii) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$

**Probability mass function (pmf)** $\leftarrow$ Discrete case

$$p(x) = P(X = x), \qquad \text{where} \sum_x p(x) = 1$$

$$F(x) = \sum_{k\leq x} p(k)$$

**Probability density function (pdf)** $\leftarrow$ Continuous case

$$f(x) = \frac{dF(x)}{dx}, \qquad \text{where} \int_{-\infty}^{\infty} f(x)dx = 1$$

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

## 2.2 Joint r.v.s

We only consider two r.v.s $X$ and $Y$.

**Joint pmf:** $\leftarrow$ Discrete case

$$p(x, y) = P\left(X = x, Y = y\right),$$

$$\sum_{\forall x} \sum_{\forall y} P\left(X = x, Y = y\right) = 1$$

**Joint pdf:** $\leftarrow$ Continuous case

$$f\left(x, y\right) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

$$\int_{\forall x} \int_{\forall y} f\left(x, y\right) dy dx = 1.$$

**Joint cdf:**

$$F(x, y) = P(X \leq x, Y \leq y)$$

$$\lim_{y \to \infty} F(x, y) = F_X(x)$$

$$\lim_{x \to \infty} F(x, y) = F_Y(y)$$

(1) $X$ and $Y$ are both discrete: joint discrete r.v.

**Joint pmf $\to$ Marginal pmf**:

$$\sum_{\forall y} p(x,y) = \sum_{\forall y} P\left(X = x, Y = y\right) = P(X = x) = p_X(x)$$

$$\sum_{\forall x} p(x,y) = \sum_{\forall x} P\left(X = x, Y = y\right) = P(Y = y) = p_Y(y)$$

**注意：要求 $X$ 的 marginal pmf，就对所有的 y 求 summation，消除 $Y$ 的 information**

**Independence:**

$X$ and $Y$ are **independent**, i.e., $X \perp\!\!\!\perp Y$, if

$$p(x,y) = p_X(x)p_Y(y), \qquad \forall x, y$$

(2) $X$ and $Y$ are both continuous: joint continuous r.v.

**Joint pdf $\to$ Marginal pdf**:

$$f_X\left(x\right) = \int_{\forall y} f\left(x,y\right) dy$$

$$f_Y\left(y\right) = \int_{\forall x} f\left(x,y\right) dx$$

**注意：要求 $X$ 的 marginal pdf，就对所有的 y 求 integral，消除 $Y$ 的信息**

**Independence:**

$X$ and $Y$ are **independent**, i.e., $X \perp\!\!\!\perp Y$, if

$$f\left(x,y\right) = f_X\left(x\right) f_Y\left(y\right), \qquad \forall x, y$$

# 3 Expectation

## 3.1 Expectation and Variance

(1) Discrete r.v. X: all possible values: $\{x_1, x_2, \ldots\}$

$$\begin{aligned} E(X) &= \sum_{\forall x_i} x_i P(X = x_i) = \sum_{\forall x_i} x_i p(x_i), \\ E(g(X)) &= \sum_{\forall x_i} g(x_i) P(X = x_i) = \sum_{\forall x_i} g(x_i) p(x_i). \end{aligned}$$

When $g(X) = X^2$,

$$E\left(X^2\right) = \sum_{\forall x_i} x_i^2 P(X = x_i) = \sum_{\forall x_i} x_i^2 p(x_i),$$

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

(2) Continuous r.v. X: with pdf $f(x)$

Idea:

$$\textit{Discrete} \to \textit{Continuous} \begin{cases} pmf \to pdf \\ \sum_{\forall x} \square \to \int_{\forall x} \square dx. \end{cases}$$

$$\begin{aligned} E(X) &= \int_{\forall x} x f(x)\, dx, \\ E(g(X)) &= \int_{\forall x} g(x) f(x)\, dx \end{aligned}$$

Similarly, if $g(X) = X^2$, we can obtain

$$E\left(X^2\right) = \int_{\forall x} x^2 f(x)\, dx.$$

Therefore, variance:

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

(3) Properties.

- $E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E\left(X_i\right)$
- $\operatorname{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}\left(X_i\right) + \left\{ \begin{array}{l} \sum_{i \neq j} a_i a_j \operatorname{Cov}\left(X_i, X_j\right) \\ 2\sum_{i<j} a_i a_j \operatorname{Cov}\left(X_i, X_j\right) \end{array} \right.$ .

## 3.2 Joint Expectation and Covariance

(1) Discrete r.v. X,Y:

$$E\left(h\left(X,Y\right)\right)=\sum_{\forall x}\sum_{\forall y}h\left(x,y\right)P\left(X=x,Y=y\right)=\sum_{\forall x}\sum_{\forall y}h\left(x,y\right)p(x,y).$$

(2) Continuous r.v. X,Y:

$$E\left(h\left(X,Y\right)\right)=\int_{\forall x}\int_{\forall y}h\left(x,y\right)f\left(x,y\right)dydx.$$

(3) Properties:

- If $X$ and $Y$ are **independent**,

  $$E\left(g_1\left(X\right)g_2\left(Y\right)\right)=E\left(g_1\left(X\right)\right)E\left(g_2\left(Y\right)\right),$$

  Therefore we can see, if $X$ and $Y$ are independent,

  $$\mathrm{Cov}\left(X,Y\right)=E\left(XY\right)-E\left(X\right)E\left(Y\right)=0.$$

- $\mathrm{Cov}\left(\sum_{i=1}^{n}a_iX_i,\sum_{j=1}^{m}b_jY_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}a_ib_j\mathrm{Cov}(X_i,Y_j)$

# 4 Review of Basic Distributions

- **Discrete Uniform**: $W_n \sim unif\{1, 2, \ldots, n\}$ or in finite scheme $W_n \sim \begin{pmatrix} 1 & \ldots & n \\ 1/n & \ldots & 1/n \end{pmatrix}$
  - $E(W_n) = \dfrac{n+1}{2}$
  - $var(W_n) = \dfrac{n^2-1}{12}$

- **Continuous Uniform**: $U \sim unif[0, 1]$
  - $1 - U \stackrel{d}{=} U$
  - CDF: $P(U \leq u) = u$
  - $E(U^k) = \frac{1}{k+1}$
  - $E(U) = \frac{1}{2}$
  - $var(U) = \frac{1}{12}$

- **Bernoulli**: $Z \sim bern(p)$ or in finite scheme $Z \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$, where $q = 1 - p$
  - $E(Z) = p$
  - $var(Z) = p - p^2 = pq$

The **gamma function** is defined as follows:

$$\Gamma(p) = \int_0^\infty z^{p-1} e^{-z} dz, \qquad \forall z \in \mathbb{R}$$

Here are some important results about $\Gamma(p)$

- For any $p > 0$,
$$\frac{\Gamma(p+1)}{\Gamma(p)} = p$$
- $\Gamma(1/2) = \sqrt{\pi}$
- If $p$ is an integer, then
$$\Gamma(p) = (p-1)!$$

- **standard gamma**: $Z_p \sim G(p)$
  - PDF: $g_p(z) = \dfrac{z^{p-1}e^{-z}}{\Gamma(p)}, \quad z > 0$
  - Scaled gamma: $X \sim G(p, \theta)$ iff $X \stackrel{d}{=} \theta Z_p$
    * $X \sim G(p_1, \theta), Y \sim G(p_2, \theta)$ and $X$ is independent of $Y$, then
$$X + Y \sim G(p_1 + p_2, \theta)$$
  - Standard exponential: If $p = 1$, $Z_1 \sim exp(1) \equiv G(1)$
  - scaled exponential: $X \sim exp(\theta) \equiv G(1, \theta)$ iff $X \stackrel{d}{=} \theta Z_1$
    * CDF: If $X \sim exp(\theta)$, then
$$P(X \leq x) = 1 - e^{-x/\theta}$$

  - $E(Z_p^s) = \dfrac{\Gamma(p+s)}{\Gamma(p)} \Rightarrow E(Z_p) = var(Z_p) = p$
  - $E(X^s) = \dfrac{\Gamma(p+s)}{\Gamma(p)}\theta^s$

- **standard normal**: $Z \sim N(0,1)$
  - $-Z \overset{d}{=} Z$
  - PDF: $\phi(z) = \dfrac{1}{\sqrt{2\pi}} e^{-z^2/2}$
  - $E(Z^n) = \begin{cases} 0, & n = 2k+1 \ (odd) \\ (2k-1)(2k-3)\cdots 1, & n = 2k \ (even) \end{cases}$
  - $Z^2/2 \sim G(1/2)$
  - location-scale normal: $X \sim N(\mu, \sigma^2)$ iff $X \overset{d}{=} \mu + \sigma Z$
    * If $X_i \overset{ind}{\sim} N(\mu_i, \sigma_i^2)$, then $\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$

- **chi-squared**: $Y \sim chisq(m) \equiv \chi^2_{(m)}$
  - $Y \overset{d}{=} 2Z$, where $Z \sim G(m/2)$
  - If $Z_i \overset{iid}{\sim} N(0,1)$, then $\sum_{i=1}^n Z_i^2 \sim \chi^2_{(n)}$

- **fisher's F**: $W \sim F(m_1, m_2)$
  - $W \overset{d}{=} \dfrac{Y_1/m_1}{Y_2/m_2}$, where $Y_i \overset{ind}{\sim} \chi^2_{(m_i)}$

- **student's t**: $T \sim t_{(m)}$
  - $T \overset{d}{=} \dfrac{\sqrt{m} Z}{\sqrt{Y}}$, where $Z \sim N(0,1), Y \sim \chi^2_{(m)}$ and $Y, Z$ are independent

## 5 Exercises (Past Exam)

**Problem 1 (Q1A)** For the random variable $X$ and $Y$, we are given that

$$EX = 4 = var(X), Y \overset{d}{=} X^2, \rho(X,Y) = 1/2 \text{ and } var(X-Y) = 3.$$

a) $E(X^2 + Y^2)$

b) $E(X+Y)^2$

**Problem 2 (Q1B)** Suppose that $U_1, U_2$ IID $U \sim \text{unif}[0,1]$ and in each case below, determine the value of $k$.

a) $P(U_1 + U_2 \leq 5/4) = k/32$

b) $P(U_1 + 2U_2 \leq 5/4) = k/32$

c) $P(U_1^2 + U_2^2 \leq 5/4) = 1/2 + k\pi$.

**Problem 3 (Q2A)** Suppose $W \sim G(2, 4)$ and evaluate the following

a) $EW$ and $\sigma(W)$

b) $P(W > 8)$

c) $P(W < 16 | W > 8)$

**Problem 4 (Q2B)** Suppose $X_i \overset{indep}{\sim} G(i, i), i = 1, \ldots, 6$, and let $V = X_6 - X_4 + X_2, W = X_5 - X_3 + X_1$.

a) Determine the mean and standard deviation for each of $V$ and $W$.

b) What is the correlation coefficient $\rho(V - W, \bar{X})$?

**Problem 5 (Q3A)** Gamma/Gauss-Maxwell: Let $W \sim G(1/2)$ and $Z \sim N(0,1)$. Starting only from the definitions based on probability density function (pdf's)

$$defn : Z \sim G(p), p > 0 \text{ iff } \mathrm{pdf}_Z = g(z) = \frac{z^{p-1}e^{-z}}{\Gamma(p)}, z > 0$$

$$defn : Z \sim N(0,1) \text{ iff } \mathrm{pdf}_Z = \phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}, -\infty < z < \infty$$

a) Prove $Z \sim N(0,1) \Leftrightarrow -Z \stackrel{d}{=} Z \;\;\&\;\; Z^2/2 \stackrel{d}{=} W$

b) Using a), or otherwise, for $Z \sim N(0,1)$, obtain $E|Z|^n, n \in \mathbb{N}$.