

Introduction to Probability

CST Part IA Paper 1

Victor Zhao
xz398@cantab.ac.uk

1 Prerequisites and Introduction

1. Combinatorics:

Counting tasks on n objects			
Permutations (sort objects)		Combinations (choose r objects)	
Distinct	Indistinct	Distinct 1 group	Distinct k groups
$n!$	$\frac{n!}{n_1!n_2!\cdots n_r!}$	$\binom{n}{r} = \frac{n!}{r!(n-r)!}$	$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1!n_2!\cdots n_k!}$

Pascal's identity: $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r} \quad (1 \leq r \leq n)$

Binomial theorem: $(x+y)^n = \sum_{r=0}^n \binom{n}{r} x^r y^{n-r}$

2. Probability axioms:

Axiom 1: For any event E , $0 \leq \mathbb{P}[E] \leq 1$

Axiom 2: Probability of the sample space S is $\mathbb{P}[S] = 1$

Axiom 3: If E and F are mutually exclusive (i.e., $E \cap F = \emptyset$), then $\mathbb{P}[E \cup F] = \mathbb{P}[E] + \mathbb{P}[F]$

In general, for all mutually exclusive events E_1, E_2, \dots ,

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} E_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[E_i]$$

3. General inclusion-exclusion principle: $\mathbb{P}\left[\bigcup_{i=1}^n E_i\right] = \sum_{r=1}^n (-1)^{r+1} \left(\sum_{i_1 < \dots < i_r} \mathbb{P}[E_{i_1} \cap \dots \cap E_{i_r}] \right)$

Case $n = 2$: $\mathbb{P}[E \cup F] = \mathbb{P}[E] + \mathbb{P}[F] - \mathbb{P}[E \cap F]$

4. Union bound (Boole's inequality): For any events E_1, E_2, \dots, E_n ,

$$\mathbb{P}\left[\bigcup_{i=1}^n E_i\right] \leq \sum_{i=1}^n \mathbb{P}[E_i]$$

5. Conditional probability (original and conditioning on event G):

Chain rule:

$$\mathbb{P}[EF] = \mathbb{P}[E|F]\mathbb{P}[F]$$

$$\mathbb{P}[EF|G] = \mathbb{P}[E|FG]\mathbb{P}[F|G]$$

Multiplication rule:

$$\mathbb{P}[E_1 E_2 \cdots E_n] = \mathbb{P}[E_1] \mathbb{P}[E_2|E_1] \cdots \mathbb{P}[E_n|E_1 \cdots E_{n-1}]$$

$$\mathbb{P}[E_1 E_2 \cdots E_n | G] = \mathbb{P}[E_1 | G] \mathbb{P}[E_2 | E_1 G] \cdots \mathbb{P}[E_n | E_1 \cdots E_{n-1} G]$$

Independence of E and F :

$$\mathbb{P}[EF] = \mathbb{P}[E]\mathbb{P}[F]$$

$$\mathbb{P}[E|F] = \mathbb{P}[E]$$

$$\mathbb{P}[EF|G] = \mathbb{P}[E|G]\mathbb{P}[F|G]$$

$$\mathbb{P}[E|FG] = \mathbb{P}[E|G]$$

Law of total probability:

$$\begin{aligned}\mathbb{P}[E] &= \mathbb{P}[EF] + \mathbb{P}[EF^c] = \mathbb{P}[E|F]\mathbb{P}[F] + \mathbb{P}[E|F^c]\mathbb{P}[F^c] \\ \mathbb{P}[E|G] &= \mathbb{P}[EF|G] + \mathbb{P}[EF^c|G] = \mathbb{P}[E|FG]\mathbb{P}[F|G] + \mathbb{P}[E|F^cG]\mathbb{P}[F^c|G]\end{aligned}$$

In general, for disjoint events F_1, F_2, \dots, F_n such that $F_1 \cup \dots \cup F_n = S$,

$$\mathbb{P}[E] = \sum_{i=1}^n \mathbb{P}[E|F_i]\mathbb{P}[F_i] \qquad \mathbb{P}[E|G] = \sum_{i=1}^n \mathbb{P}[E|F_iG]\mathbb{P}[F_i|G]$$

Bayes' theorem:

$$\mathbb{P}[F|E] = \frac{\mathbb{P}[E|F]\mathbb{P}[F]}{\mathbb{P}[E]} \qquad \mathbb{P}[F|EG] = \frac{\mathbb{P}[E|FG]\mathbb{P}[F|G]}{\mathbb{P}[E|G]}$$

6. Confusion matrix:

		Actual condition	
		Positive F	Negative F^c
Predicted condition	Positive E	True positive $\mathbb{P}[E F]$	False positive $\mathbb{P}[E F^c]$
	Negative E^c	False negative $\mathbb{P}[E^c F]$	True negative $\mathbb{P}[E^c F^c]$

2 Random Variables

1. Probability distribution functions:

Discrete random variable X :

- Probability mass function (PMF): $p(x)$
- Compute probability:

$$\begin{aligned}\mathbb{P}[X = a] &= p(a) \\ \mathbb{P}[a \leq X \leq b] &= \sum_{x=a}^b p(x)\end{aligned}$$

- Cumulative distribution function (CDF):

$$F_X(a) = \mathbb{P}[X \leq a] = \sum_{x \leq a} p(x)$$

Continuous random variable X :

- Probability density function (PDF): $f(x)$
- Compute probability:

$$\begin{aligned}\mathbb{P}[X = a] &= 0 \\ \mathbb{P}[a \leq X \leq b] &= \int_a^b f(x)dx\end{aligned}$$

- Cumulative distribution function (CDF):

$$F_X(a) = \mathbb{P}[X \leq a] = \int_{-\infty}^a f(x)dx$$

2. Expectation:

Discrete random variable X :

$$\begin{aligned}\mathbb{E}[X] &= \sum_x xp(x) \\ \mathbb{E}[g(X)] &= \sum_x g(x)p(x)\end{aligned}$$

Continuous random variable X :

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} xf(x)dx \\ \mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} g(x)f(x)dx\end{aligned}$$

Linearity of expectation: $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

Additivity of expectation: $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

3. Variance: $\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Scaling of variance: $\mathbb{V}[aX + b] = a^2\mathbb{V}[X]$

Standard deviation: $\mathbb{SD}[X] = \sqrt{\mathbb{V}[X]}$

Scaling of standard deviation: $\mathbb{SD}[aX + b] = |a|\mathbb{SD}[X]$

4. Discrete distributions:

Bernoulli $\text{Ber}(p)$: 1 experiment with success probability p

$$\mathbb{P}[X = 1] = p \quad \mathbb{E}[X] = p \quad \mathbb{V}[X] = p(1 - p)$$

Binomial $\text{Bin}(n, p)$: n independent trials with success probability p

$$\mathbb{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \quad \mathbb{E}[X] = np \quad \mathbb{V}[X] = np(1 - p)$$

Poisson $\text{Pois}(\lambda)$: # successes over experiment duration, with success rate $\lambda = np$

$$\mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda} \quad \mathbb{E}[X] = \lambda \quad \mathbb{V}[X] = \lambda$$

Geometric $\text{Geo}(p)$: # independent trials until first success, with success probability p

$$\mathbb{P}[X = n] = (1 - p)^{n-1} p \quad \mathbb{E}[X] = \frac{1}{p} \quad \mathbb{V}[X] = \frac{1 - p}{p^2}$$

Negative binomial $\text{NegBin}(r, p)$: # independent trials until r success, with success probability p

$$\mathbb{P}[X = n] = \binom{n-1}{r-1} (1 - p)^{n-r} p^r \quad \mathbb{E}[X] = \frac{r}{p} \quad \mathbb{V}[X] = \frac{r(1 - p)}{p^2}$$

Hypergeometric $\text{Hyp}(N, n, m)$: # objects with a feature in a sample of size n (without replacement) from a population of size N that contains m items with the feature

$$\mathbb{P}[X = n] = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}} \quad \mathbb{E}[X] = n \frac{m}{N} \quad \mathbb{V}[X] = n \frac{m}{N} \left(1 - \frac{m}{N}\right) \left(1 - \frac{n-1}{N-1}\right)$$

5. Continuous distributions:

Uniform $\text{Uni}(\alpha, \beta)$: equal probability within range $[\alpha, \beta]$

$$\begin{aligned} \text{PDF: } f(x) &= \begin{cases} \frac{1}{\beta - \alpha} & \text{when } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases} & \text{CDF: } F(x) &= \begin{cases} 0 & \text{when } x < \alpha \\ \frac{x - \alpha}{\beta - \alpha} & \text{when } \alpha \leq x \leq \beta \\ 1 & \text{when } x > \beta \end{cases} \\ \mathbb{E}[X] &= \frac{\alpha + \beta}{2} & \mathbb{V}[X] &= \frac{(\beta - \alpha)^2}{12} \end{aligned}$$

Exponential $\text{Exp}(\lambda)$: time until first success occurs, with success rate λ

$$\begin{aligned} \text{PDF: } f(x) &= \begin{cases} \lambda e^{-\lambda x} & \text{when } x \geq 0 \\ 0 & \text{otherwise} \end{cases} & \text{CDF: } F(x) &= 1 - e^{-\lambda x} \\ \mathbb{E}[X] &= \frac{1}{\lambda} & \mathbb{V}[X] &= \frac{1}{\lambda^2} \end{aligned}$$

Normal (Gaussian) $\mathcal{N}(\mu, \sigma^2)$: mean μ , variance σ^2

$$\begin{aligned} \text{PDF: } f(x) &= \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}} & \text{CDF: } F(x) &= \Phi\left(\frac{x - \mu}{\sigma}\right) \\ \mathbb{E}[X] &= \mu & \mathbb{V}[X] &= \sigma^2 \end{aligned}$$

$$X \sim \mathcal{N}(\mu, \sigma^2) \implies aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2), Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \implies X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

6. Continuity correction:

Discrete		Continuous
$\mathbb{P}[X = a]$	\approx	$\mathbb{P}[a - 0.5 \leq X \leq a + 0.5]$
$\mathbb{P}[X > a]$	\approx	$\mathbb{P}[X \geq a + 0.5]$
$\mathbb{P}[X \geq a]$	\approx	$\mathbb{P}[X \geq a - 0.5]$
$\mathbb{P}[X < a]$	\approx	$\mathbb{P}[X \leq a - 0.5]$
$\mathbb{P}[X \leq a]$	\approx	$\mathbb{P}[X \leq a + 0.5]$

7. Joint probability mass function (for discrete RVs): $p_{X,Y}(a, b) = \mathbb{P}[X = a, Y = b]$

Joint distribution function (for discrete or continuous RVs): $F_{X,Y}(a, b) = \mathbb{P}[X \leq a, Y \leq b]$

Joint probability density f and joint continuous distribution F (for continuous RVs):

$$F(a, b) = \int_{-\infty}^a \int_{-\infty}^b f(x, y) dx dy \quad f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

$$\mathbb{P}[a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2] = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y) dx dy$$

Marginal distribution: $F_X(a) = \mathbb{P}[X \leq a] = \lim_{b \rightarrow \infty} F_{X,Y}(a, b)$

8. Covariance: $\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Covariance of linear combinations:

$$\begin{aligned} \text{Cov}[X, a] &= 0 & \text{Cov}[X, X] &= \mathbb{V}[X] \\ \text{Cov}[aX, bY] &= ab\text{Cov}[X, Y] & \text{Cov}[X + a, Y + b] &= \text{Cov}[X, Y] \end{aligned}$$

Variance of sum: $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}[X, Y]$

In general, for any random variables X_1, X_2, \dots, X_n :

$$\mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{V}[X_i] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}[X_i, X_j]$$

Correlation coefficient: $\rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}} \in [-1, 1]$ ($\rho(X, Y) = 0$ if $\mathbb{V}[X] = 0$ or $\mathbb{V}[Y] = 0$)

Scaling-invariance of correlation coefficient: $\rho(aX, bY) = \rho(X, Y)$

3 Moments and Limit Theorems

1. Markov's inequality: for any non-negative random variable X with finite $\mathbb{E}[X]$, for any $a > 0$,

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

Let $a = \delta \cdot \mathbb{E}[X]$ (where $\delta > 0$), then the inequality can be rewritten as

$$\mathbb{P}[X \geq \delta \cdot \mathbb{E}[X]] \leq \frac{1}{\delta}$$

2. Chebyshev's inequality: for any random variable X with finite $\mathbb{E}[X]$ and $\mathbb{V}[X]$, for any $a > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq a] \leq \frac{\mathbb{V}[X]}{a^2}$$

Let $a = \sqrt{\delta \cdot \mathbb{V}[X]}$ (where $\delta > 0$), then the inequality can be rewritten as

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \sqrt{\delta \cdot \mathbb{V}[X]}\right] \leq \frac{1}{\delta}$$

3. Weak law of large numbers: let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, where X_i 's are independent and identically distributed (i.i.d.) with finite expectation μ and finite variance σ^2 . Then, for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} [|\bar{X}_n - \mu| > \epsilon] = 0$$

Strong law of large numbers:

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \bar{X}_n = \mu \right] = 1$$

4. Central limit theorem: let X_1, X_2, \dots, X_n be any sequence of i.i.d. random variables with finite expectation μ and finite variance σ^2 . Let

$$Z_n = \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} = \frac{1}{\sigma\sqrt{n}} \left(\sum_{i=1}^n X_i - n\mu \right)$$

Then for any number $a \in \mathbb{R}$, it holds that

$$\lim_{n \rightarrow \infty} F_{Z_n}(a) = \Phi(a) = \frac{1}{2\pi} \int_{-\infty}^a e^{-\frac{x^2}{2}} dx$$

where Φ is the CDF of the standard normal distribution $\mathcal{N}(0, 1)$.

4 Applications and Statistics

- 1.