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
$n!$	$\dfrac{n!}{n_1!n_2!\cdots n_r!}$	$\dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$	$\dbinom{n}{n_1, n_2, \cdots, n_k} = \dfrac{n!}{n_1!n_2!\cdots n_k!}$

Pascal's identity: $\dbinom{n}{r} = \dbinom{n-1}{r-1} + \dbinom{n-1}{r}$ $(1 \le r \le n)$

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

2. Probability axioms:

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

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

   Axiom 3: If E and F are mutually exclusive (i.e., $E \cap F = \varnothing$), then $\mathbb{P}[E \cup F] = \mathbb{P}[E] + \mathbb{P}[F]$
   In general, for all mutually exclusive events $E_1, E_2, \cdots$,

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

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

   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 $E_1, E_2, \cdots, E_n$,

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

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

   Chain rule:

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

   Multiplication rule:

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

   Independence of E and F:

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

Law of total probability:

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

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

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

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

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	
	Total population	Positive F	Negative $F^{\complement}$
Predicted condition	Positive E	True positive $\mathbb{P}[E|F]$	False positive $\mathbb{P}[E|F^{\complement}]$
	Negative $E^{\complement}$	False negative $\mathbb{P}[E^{\complement}|F]$	True negative $\mathbb{P}[E^{\complement}|F^{\complement}]$

2   Random Variables

1. Probability distribution functions:

Discrete random variable X:

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

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

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

- 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:

$$\mathbb{P}[X = a] = 0$$

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

- Cumulative distribution function (CDF):

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

2. Expectation:

Discrete random variable X:

$$\mathbb{E}[X] = \sum_{x} xp(x)$$

$$\mathbb{E}[g(X)] = \sum_{x} g(x)p(x)$$

Continuous random variable X:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx$$

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

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: $\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$

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

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

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

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

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

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

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

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

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

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

Hypergeometric $\mathrm{Hyp}(N,n,m)$: # 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

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

5. Continuous distributions:

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

$$\text{PDF: } f(x)=\begin{cases}\frac{1}{\beta-\alpha} & \text{when } \alpha\le x\le\beta \\ 0 & \text{otherwise}\end{cases} \qquad\qquad \text{CDF: } F(x)=\begin{cases}0 & \text{when } x<\alpha \\ \frac{x-\alpha}{\beta-\alpha} & \text{when } \alpha\le x\le\beta \\ 1 & \text{when } x>\beta\end{cases}$$

$$\mathbb{E}[X]=\frac{\alpha+\beta}{2} \qquad\qquad\qquad \mathbb{V}[X]=\frac{(\beta-\alpha)^2}{12}$$

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

$$\text{PDF: } f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{when } x\ge 0 \\ 0 & \text{otherwise}\end{cases} \qquad\qquad \text{CDF: } F(x)=1-e^{-\lambda x}$$

$$\mathbb{E}[X]=\frac{1}{\lambda} \qquad\qquad\qquad \mathbb{V}[X]=\frac{1}{\lambda^2}$$

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

$$\text{PDF: } f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu^2)}{2\sigma^2}} \qquad\qquad \text{CDF: } F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$\mathbb{E}[X]=\mu \qquad\qquad\qquad \mathbb{V}[X]=\sigma^2$$

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

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

6. 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): $F_{X,Y}(a,b) = \mathbb{P}[X \le a, Y \le b]$

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

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

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

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

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

Covariance of linear combinations:

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

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

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

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

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

4 Applications and Statistics
1.