ST2131 Finals Cheatsheet by Wang Xiuxuan

1 Chapter 1

2 Chapter 2

3 Chapter 3

4 Chapter 4

4.1 Random Variables

Definition 4.1: A **random variable**, X, is a mapping from the sample space to real numbers.

4.2 Discrete Random Variables

Definition 4.2: A random variable is said to be **discrete** if range of X is either **fininite** or **countably infinite**.
Definition 4.3: The **probability mass function** (pmf) of a discrete random variable X is defined by

$$p_X(x) = \begin{cases} P(X=x), & \text{if } x = x_1, x_2, \cdots \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\sum_{i=1}^{\infty} p_X(x_i) = 1$.
Definition 4.4: The **cumulative distribution function** (cdf) of a discrete random variable X is defined as $F_X : \mathbb{R} \to \mathbb{R}T$ where $F_X(x) = P(X \leq x)$, for $x \in \mathbb{R}$.

4.3 Expected Value

Definition 4.5: If X is a discrete random variable having a probability mass function p_X, the **expectation** or **expected value** of X is defined by

$$E(X) = \sum_x x p_X(x).$$

Tail Sum Formula: For **nonnegative integer-valued** random variable X, we have

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k).$$

4.4 Expectation of a Function of a Random Variable

Proposition 4.1 If X is a **discrete random variable** that takes values x_i, $i \geq 1$, with respective probabilities $p_X(x_i)$, then for any real value function g,

$$\begin{aligned} E[g(X)] &= \sum_i g(x_i) p_X(x_i) \quad \text{or equivalently} \\ &= \sum_x g(x) p_X(x). \end{aligned}$$

Corollary 4.2 Let a and b be constants, then

$$E[aX+b] = aE(X) + b.$$

4.5 Variance and Standard Deviation

Definition 4.6 If X is a random variable with mean μ, then the **variance** of X, denoted by $\mathrm{Var}(X)$, is defined by

$$\mathrm{Var}(X) = E[(X-\mu)^2].$$

Definition 4.7 The **standard deviation** of X, denoted by σ_X or $\mathrm{SD}(X)$, is defined by

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Scaling and shifting:
- $\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$.
- $\mathrm{SD}(aX+b) = |a| \mathrm{SD}(X)$.

4.6 Discrete Random Variables Arising from Repeated Trials

Bernoulli Random Variable: $\mathrm{Be}(p)$, define

$$X = \begin{cases} 1, & \text{if it is a success;} \\ 0 & \text{if it is a failure.} \end{cases}$$

- $P(X=1) = p$;
- $E(X) = p$;
- $P(X=0) = 1-p$;
- $\mathrm{Var}(X) = p(1-p)$.

Binomial Random Variable: $\mathrm{Bin}(n,p)$, define X as **number of successes** in n $\mathrm{Bernoulli}(p)$ trials.
Therefore, X takes values $0, 1, 2, \cdots, n$.
- $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$;
- $E(X) = np$;
- $\mathrm{Var}(X) = np(1-p)$;

Geometric Random Variable: $\mathrm{Geom}(p)$, define X as **number of trials** until the first success in a sequence of independent $\mathrm{Bernoulli}(p)$ trials.
- $P(X=k) = pq^{k-1}$;
- $E(X) = \frac{1}{p}$;
- $\mathrm{Var}(X) = \frac{1-p}{p^2}$;

Alternative Geometric Distribution: Let X' be the **number of failures** in the $\mathrm{Bernoulli}(p)$ trials in order to obtain the first success. Here,

$$X = X' + 1$$

- $P(X'=k) = pq^k$;
- $E(X') = \frac{1-p}{p}$;
- $\mathrm{Var}(X') = \frac{1-p}{p^2}$;

Negative Binomial Random Variable: $\mathrm{NB}(r,p)$, define X to be **number of** $\mathrm{Bernoulli}(p)$ **trials** required to obtain r **successes**. For $k \geq r$, we have:
- $P(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$;
- $E(X) = \frac{r}{p}$;
- $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$;

Remark: $\mathrm{Geom}(p) = \mathrm{NB}(1,p)$.

4.7 Poisson Random Variable

Poisson: A random variable X is said to have a **Poisson** distribution with parameter λ if X takes values $0, 1, 2, \cdots$ with probabilities given as:

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{for } k = 0, 1, 2, \cdots$$

- $E(X) = \lambda$;
- $\mathrm{Var}(X) = \lambda$.

When n is large and λ is moderate, we can use **poisson** to approximate **binomial**:

$$\mathrm{Bin}(n,p) \approx \mathrm{Poisson}(np).$$

I.e. $P(X=k) \approx e^{-\lambda}\frac{\lambda^k}{k!}$.
Examples obeying Poisson distribution:
- Number of misprints on a page;
- Number of people in a community living to 100 years;
- Number of wrong telephone numbers that are dialled in a day;
- Number of people entering a store on a given day;
- Number of particles emitted by a radioactive source.

These are approximately Poisson **because of the Poisson approximation to the binomial distribution.**