ST2131 Cheatsheet
by Yiyang, AY20/21

Chapter 01 - Combinatorial Analysis

Some Combinatorial Identities

For all non-negative integers m, n, k and $k \leq n$,

- $k\binom{n}{k} = (n-k+1)\binom{n}{k-1} = n\binom{n-1}{k-1}$ *(AY20/21Sem2 Tut1Qn7)*
- $\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}$ *(AY20/21Sem2 Tut1Qn8)*
- $\binom{n+m}{k} = \binom{n}{0}\binom{m}{k} + \binom{n}{1}\binom{m}{k-1} + \ldots + \binom{n}{r}\binom{m}{0}$ *(AY20/21Sem2 Tut1Qn9)*
- $\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2$ *(AY20/21Sem2 Tut1Qn10)*

Chapter 02 - Axioms of Probability

Inclusion/Exclusion Principle (Prop. 2.6)

Let $E_1, E_2, \ldots, E_n$ be any n events, then

$$P(E_1 \cup E_2 \cup \ldots E_n) = \sum_{r=1}^{n} \left((-1)^{r+1} \sum_{1 \leq i_1 < \ldots < i_r \leq n} P(E_{i_1} \cap \ldots \cap E_{i_r})\right)$$

Tut2Qn16 - Generalised Bonferroni's Inequality

Let $E_1, E_2, \ldots, E_n$ be any n events, then

$$P(E_1 E_2 \ldots E_n) \geq P(E_1) + \ldots + P(E_n) - (n-1)$$

When $n = 2$,

$$P(E_1 E_2) \geq P(E_1) + P(E_2) - 1$$

Chapter 03 - Conditional Probability

Bayes' Formulae

Suppose events A_1, A_2, ..., A_n partition the sample space, and $P(A_i) > 0$, for all $i = 1, 2, \ldots, n$.
Then for any event B, and any $1 \leq i \leq n$,

$$P(B) = \sum_{j=1}^{n} P(B|A_j)P(A_j) \tag{1}$$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)} \tag{2}$$

Some Identities Involving Conditional Probability

For any events A, B, C,

$$\begin{aligned} P(A|C) &= P(AB|C) + P(AB^C|C) \\ &= P(A|BC)P(B|C) + P(A|B^C C)P(B^C|C) \end{aligned}$$

Chapter 04 - Random Variables

Tail Sum Formula

For *non-negative integer-valued* random variable X, if X is a D.R.V. (i.e. $X = 0, 1, \ldots, 2$),

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

or if X is a C.R.V.,

$$E(X) = \int_0^{\infty} P(X > x)\, dx = \int_0^{\infty} P(X \geq x)\, dx$$

Chapter 05 - Continuous Random Variable

Distribution of a Function of R.V.

For r.v. X with pdf. $f_X(x)$, assume $g(x)$ is a function of X that is **strictly monotonic** and **differentiable**. Then the pdf. of $Y = g(X)$,

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)|, & y = g(x) \text{ for some } x \\ 0, & \text{otherwise} \end{cases}$$

Binomial to Normal Approximation

(Remember **Continuity Correction**!!!)
For $X \sim Bin(n, p)$ where npq is large (generally good when $npq \geq 10$),

$$Bin(n, p) \approx N(np, npq) \text{ , i.e. } \frac{X - np}{\sqrt{npq}} \approx Z$$

Binomial to Poisson Approximation

For $X \sim Bin(n, p)$ where n is large and p (or q) is small so that np (or nq) is moderate.

- when $p < 0.1$, $Bin(n, p) \approx Poisson(np)$
- when $p > 0.9$, $Bin(n, q) \approx Poisson(nq)$

Chapter 06 - Joint Distributions

Convolution of Independent Distributions

$$F_{X+Y}(a) = \int_{-\infty}^{\infty} F_Y(a-x)\, f_X(x)\, dx = \int_{-\infty}^{\infty} F_X(a-y)\, f_Y(y)\, dy$$

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_Y(a-x)\, f_X(x)\, dx = \int_{-\infty}^{\infty} f_X(a-y)\, f_Y(y)\, dy$$

Prop.6.4 - Sum of Independent Gamma R.V.s

Assume $X \sim Gamma(\alpha, \lambda)$ and $Y \sim Gamma(\beta, \lambda)$ are independent.

$$X + Y \sim Gamma(\alpha + \beta, \lambda)$$

Prop.6.5 - Sum of Independent Normal R.V.s

Assume $X_i, i = 1, 2, \ldots, n$ are independent random variables that are normally distributed with parameters $\mu_i, \sigma_i^2,\ i = 1, 2, \ldots, n$.

$$\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i \ \sum_{i=1}^{n} \sigma_i^2)$$

Ex.6.18 - Sum of Independent Poisson R.V.s

Assume $X \sim Poisson(\lambda),\ Y \sim Poisson(\mu)$ are independent.

$$X + Y \sim Poisson(\lambda + \mu)$$

Ex.6.19 - Sum of Independent Binomial R.V.s

Assume $X \sim Bin(n, p),\ Y \sim Bin(m, p)$ are independent.

$$X + Y \sim Bin(n + m, p)$$

Note: This statement only works when the second parameter of both R.V.s are the same (i.e. both p). For problems with different parameters and large values, can consider using Normal Approximation with (*Prop.6.5*).

Ch 07 - Properties of Expectation

Ex.7.20 - Expectation of a Random Sum

Suppose $X_1, X_2, \ldots$ are i.i.d. with common mean μ. Suppose N is a non-negative integer-valued random variable independent of the X_i.

$$\sum_{k=1}^{N} X_k = \mu E[N]$$

Common Moment Generating Functions

- $X \sim Be(p),\ M_X(t) = 1 - p + pe^t$
- $X \sim Bin(n, p),\ M_X(t) = (1 - p + pe^t)^n$
- $X \sim Geom(p),\ M_X(t) = \frac{pe^t}{1-qe^t}$
- $X \sim Poisson(\lambda),\ M_X(t) = e^{\lambda(e^t - 1)}$
- $X \sim U(\alpha, \beta),\ M_X(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t}$
- $X \sim Exp(\lambda),\ M_X(t) = \frac{\lambda}{\lambda - t}$, for $t < \lambda$
- $X \sim N(\mu, \sigma^2),\ M_X(t) = e^{(\mu t + \sigma^2 t^2/2)}$

Less Common MGFs

- (*Ex.7.31*) X is a chi-squared r.v. with n deg. of freedom, $M_X(t) = (E[e^{tZ^2}])^n = (1 - 2t)^{-n/2}$

Ex.7.34 - "Partitioned" Poisson Distribution

Let X be the r.v. that denotes total number of events. Suppose each event is a Ber. process with p probability of being A and $q = (1-p)$ being B. Let X_A, X_B denote the number of events that are A and B respectively. If $X \sim Poisson(\lambda)$,

$$X_A \sim Poisson(p\lambda),\ X_B \sim Poisson(q\lambda)$$

Ch 08 - Limit Theorems

Markov's Inequality

For **non-negative** r.v. X and any $a > 0$,

$$P(X \geq a) \leq \frac{EX}{a}$$

Chebyshev's Inequality

Let X be a r.v. with mean μ, then for any $a > 0$,

$$P(|X - \mu| \geq a) \leq \frac{\text{var}(X)}{a^2}$$

One-sided Chebyshev's Inequality

Let X be a r.v. with **zero mean** and variance σ^2, then for any $a > 0$,

$$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Central Limit Theorem

(Remember **Continuity Correction** when a CRV is used to approximate a DRV!!!) For a sequence of i.i.d. r.v.s $X_1, X_2, \ldots$, each with mean μ and variance σ^2,

$$\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \to Z, \text{ as } n \to \infty$$

WLLN SLLN

Jensen's Inequality

For any r.v. X and convex function $g(X)$,

$$E[g(X)] \geq g(E[X])$$

, provided the expectations exist and are finite.

D.R.V. Models

Bernoulli

$X \sim Be(p)$, indicate whether an event is successful.
Parameter - $p = P(X=1)$: success rate
Distribution - $P(X=1) = p,\ P(X=0) = q = 1-p$
$E(X) = p,\ \text{var}(X) = pq = p(1-p)$

Binomial

$X \sim Bin(n,p)$, total number of successes in n i.i.d. $Be(p)$ trials.
Parameters

- n : number of trials
- p : success rate for each Bernoulli trial

Distribution

$$P(X=k) = \binom{n}{k} p^x q^{n-x},\ k = 0, 1, \ldots, n$$

$E(X) = np,\ \text{var}(X) = npq = np(1-p)$

Geometric

$X \sim Geom(p)$, number of i.i.d $Be(p)$ trials until one success. $X = 1, 2, \ldots$
Parameter - p : success rate
Distribution

$$P(X=k) = pq^{k-1},\ k = 1, 2, \ldots$$

Memoryless Property: $P(X > s+t | X > s) = P(X > t)\ \ s, t > 0.$
$E(X) = \frac{1}{p},\ \text{var}(X) = \frac{1-p}{p^2}$

Negative Binomial

$X \sim NB(r,p)$, number of i.i.d $Be(p)$ trials for first r successes. $X = r, r+1, \ldots$
Parameter

- r : successes needed
- p : success rate

Distribution

$$P(X=k) = \binom{k-1}{r-1} p^r q^{x-r},\ k = r, r+1, \ldots$$

$E(X) = \frac{r}{p},\ \text{var}(X) = \frac{r(1-p)}{p^2}$
$Geom(p) = NB(1,p)$

Poisson

$X \sim Poisson(\lambda)$
Parameter - λ : "average occurrence rate in unit time interval"
Distribution

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, \ldots$$

$E(X) = \text{var}(X) = \lambda$

Hypergeometric

Suppose there are N identical balls, m of them are red and $N-m$ are blue. $X \sim H(n, N, m)$ is the number of red balls obtained in n draws without replacement.
Parameter

- N : total number of objects ("red and blue balls")
- m : number of objects considered success ("red balls")
- n : number of trials without replacement ("draws")

Distribution

$$P(X=k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}},\ k = 0, 1, \ldots, n$$

$E(X) = \frac{nm}{N},\ \text{var}(X) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right]$

C.R.V. Models

Uniform

$X \sim U(a,b)$, where X has equal probability of taking any value in (a,b).
Parameters - a and b : the start and end value for the interval
Distribution

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & b \leq x \end{cases}$$

$E(X) = \frac{a+b}{2},\ \text{var}(X) = \frac{(b-a)^2}{12}$

Exponential

$X \sim Exp(\lambda)$ usually models the life time of a product, for $\lambda > 0$
Distribution

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}$$

Memoryless Property: $P(X > s+t | X > s) = P(X > t)\ \ s, t > 0.$
$E(X) = \frac{1}{\lambda},\ \text{var}(X) = \frac{1}{\lambda^2}$

Normal

$X \sim N(\mu, \sigma^2)$. Special case : $Z \sim N(0,1)$ standard normal
Parameters

- μ : mean
- σ : standard deviation

Distribution

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},\ z \in \mathbb{R}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)},\ x \in \mathbb{R}$$

$E(X) = \mu,\ \text{var}(X) = \sigma^2$

Gamma

$X \sim Gamma(\alpha, \lambda)$ can be seen as the sum of α independent $Exp(\lambda)$, for $\alpha, \lambda > 0$. (Refer to *Prop.6.4*)
Parameters

- α : shape parameter
- λ : rate parameter
- ($\frac{1}{\lambda}$: scale parameter)

Distribution

$$f(x) = \begin{cases} \frac{\lambda e^{\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$Exp(\lambda) = Gamma(1, \lambda)$ is a special case of Gamma r.v.

$E(X) = \frac{\alpha}{\lambda}$, $\text{var}(X) = \frac{\alpha}{\lambda^2}$
Gamma Function $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1}\, dy$
It satisfies that

- $\Gamma(1) = 1,\ \Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1),\ \alpha > 0$
- $\Gamma(n) = (n-1)!,\ n \in \mathbb{Z}^+$

Weibull Distribution

$S \sim W(\nu, \alpha, \beta)$ can be seen as the generalised form of Exponential r.v.

- $E(\lambda) = W(1, \lambda, 0)$
- If $X \sim E(\lambda)$, then linear transformation $Y = \alpha X + \nu \sim W(\nu, \alpha, \lambda)$ (*Tut7Qn15*)

Cauchy

$X \sim \text{Cauchy}(\theta, \alpha)$ for $\theta \in \mathbb{R}, \alpha > 0$ if it has the distribution:

$$f(x) = \frac{1}{\pi\alpha[1 + (\frac{x-\theta}{\alpha})^2]},\ x \in \mathbb{R}$$

$E(X)$ and $\text{var}(X)$ do not exist for Cauchy r.v.

Beta

$X \sim B(a, b)$. Specifically, $U(0,1) = B(1,1)$ is a special case of Beta r.v.

Common Identities

$\sum_{i=1}^\infty ir^{i-1} = \frac{1}{(1-r)^2}$, for $|r| < 1$