MATH 318 Formula Sheet

Probability Theory

Permutations and Combinations:

$$P(n,r) = \frac{n!}{(n-r)!} \qquad C(n,r) = \binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Multinomial Coefficient:

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{m-1}}{n_m} = \frac{n!}{(n_1!)(n_2!)\cdots(n_m!)}$$

Probability Function:

- $0 \le P(E) \le 1$
- $P(S) = 1$
- $E_1 \cap E_2 = \varnothing \implies P(E_1 \cup E_2) = P(E_1) + P(E_2)$

Inclusion-Exclusion Principle:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Conditional Probability:

$$P(E \mid F) = \frac{P(E\cap F)}{P(F)}$$

Independence:

$$P(E \cap F) = P(E)P(F)$$

Law of Total Probability ($F_1, \dots, F_n$ is a partition of S):

$$P(E) = \sum_{i=1}^{n} P(E \mid F_i)P(F_i)$$

Bayes' Theorem:

$$P(F_j \mid E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

Random Variables

Probability Mass Function:

$$p(x) = P(X = x)$$

Probability Density Function:

$$P(a \le X \le b) = \int_a^b f(x)\,\mathrm{d}x$$

Cumulative Distribution Function:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(s)\,\mathrm{d}s$$

$$F'(x) = f(x)$$

Memoryless Property:

$$P(X > m+n \mid X \ge n) = P(X > m)$$

Expectation:

$$\mathbb{E}(X) = \sum_i x_i p(x_i) \qquad \mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

Linearity of Expectation:

$$\mathbb{E}(aX+b) = a\mathbb{E}(X) + b \quad \mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

Normal Distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Scaling Property:

$$X \sim N(\mu, \sigma^2) \implies Y = \frac{X-\mu}{\sigma} \sim N(0,1)$$

Law of the Unconscious Statistician:

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x$$

Moments:

$$\mathbb{E}(X^n) = \begin{cases} \int_{-\infty}^{\infty} x^n f(x)\,\mathrm{d}x \\ \sum_i x_i^n p(x_i) \end{cases}$$

Variance:

$$\mathrm{Var}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \mathbb{E}\big(X^2\big) - \mathbb{E}(X)^2$$

Joint Distribution:

$$p(x,y) = P(X = x, Y = y)$$

$$P((X,Y) \in C) = \iint_C f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

Marginal PDF:

$$P(X \in A) = \int_A \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_A f_X(x)\,\mathrm{d}x$$

Independence:

$$f(x,y) = f_X(x) f_Y(y)$$

Expectation and Variance of Independent RVs:

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Covariance:

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Correlation Coefficient:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1,1]$$

Sums of Independent RVs:

$$F_{X+Y}(a) = \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y$$

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

Poisson Process $X_i \sim \exp(\lambda)$:

$$f_{X_1+\cdots+X_n} = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$$

$$N_t = \mathrm{Pois}(\lambda t)$$

Conditional Distribution:

$$p_{X+Y}(x \mid y) = \frac{p(x,y)}{p_Y(y)} \qquad f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}$$

Conditional Expectation:

$$\mathbb{E}(X \mid Y = y) = \sum_x p_{X+Y}(x \mid y)$$

$$\mathbb{E}(X) = \sum_y \mathbb{E}(X \mid Y = y)P(Y = y) = \mathbb{E}(\mathbb{E}(X \mid Y))$$

Characteristic Functions

$$\phi(t) = \mathbb{E}\big(e^{itX}\big) \qquad M(t) = \mathbb{E}\big(e^{tX}\big)$$

Extracting Moments:

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n}\Big|_{t=0}\phi(t) = i^n \mathbb{E}(X^n)$$

Inversion Theorem:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\,\mathrm{d}t$$

Shifting Property:

$$\phi_{aX+b}(t) = e^{itb}\phi_X(ta)$$

Convergence of Random Variables

Convergence in Distribution:

$$X_n \xrightarrow{D} X \iff \lim_{n\to\infty} F_n(x) = F(x) \forall \text{cont. } x$$

Continuity Theorem:

$$F_n \to F \implies \lim_{n\to\infty} \phi_n(t) = \phi(t) \forall t \in \mathbb{R}$$

$$\lim_{n\to\infty} \phi_n(t) = \phi(t) \wedge \text{cont. at } 0 \implies X_n \xrightarrow{D} X$$

Thm (*Weak Law of Large Numbers*): let $X_1, X_2, \cdots$ be iid RVs with $\mu = \mathbb{E}(X_i) < \infty$.

$$S_n = \sum_{i=1}^{n} X_i \implies \frac{S_n}{n} \xrightarrow{D} \mu$$

Thm (*Central Limit Theorem*): let $X_1, X_2, \cdots$ be iid RVs with $\mu = \mathbb{E}(X_i)$ and $\sigma^2 = \mathrm{Var}(X_i) < \infty$.

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$$

Statistical Estimation

Sample Mean:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Sample Variance:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Hypothesis Testing:

Reject if $P(\text{observation or worse} \mid H) < p = 0.05$

$P(|\bar{X} - \mu| \ge a) = 0.05$, solve for a using CLT

Reject if observed a is greater than calculated a

a% Confidence Interval A:

$$P(\bar{X} \in A) = a\%$$

$$P(|Z| < z) = P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < z\right) = a\%$$

$$\bar{X} \in \left[\mu - \frac{\sigma}{\sqrt{n}}z, \mu + \frac{\sigma}{\sqrt{n}}z\right]$$

$$\mu \in \left[\bar{X} - \frac{\sigma}{\sqrt{n}}z, \bar{X} + \frac{\sigma}{\sqrt{n}}z\right]$$

Student-t Distribution ($n-1$ DOF):

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

$$P(|T| > a) = 0.05$$

Random Walks

Simple Random Walks on $\mathbb{Z}^d$:

- Number of visits to the origin: $\mathbb{E}(M) = \frac{1}{1-u}$
- Probability of return: $u = 1 - \frac{1}{\mathbb{E}(M)}$
- If $u = 1$, the walk is recurrent, otherwise transient.
- $\mathbb{E}(M) = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} \frac{1}{1-\phi_1(\vec{k})} d^d\vec{k}$
- SRW is recurrent for $d = 1, 2$, transient for $d > 2$.

Markov Chains

Transition Matrix P: rows add to 1.

$\vec{X}_{n+1} = \vec{X}P$

One-Step Transition Probability:

$P_{ij} = P(X_{n+1} = j \mid X_n = i) \qquad i \to j$

$\sum_j P_{ij} = 1$

n-step Transition Probability:

$P_{ij}^n = P(X_{l+n} = j \mid X_l = i)$

Chapman-Kolmogorov Equation:

$P_{ij}^{n+m} = \sum_k P_{ik}^n P_{kj}^m$

$P^{n+m} = P^n P^m$

Classification of States:

- State i is absorbing if $P_{ii} = 1$.
- j is accessible from i if $P_{ij}^n > 0$ for some n.
- i and j communicate ($i \leftrightarrow j$) if j is accessible from i and i is accessible from j.

Irreducibility: all states communicate.

Periodicity:

$d = \gcd\{n \geq 1 : P_{ii}^n > 0\}$

$d = 1$ or $P_{ii}^n = 0 \forall n \implies i$ is aperiodic

Transience and Recurrence:

$f_i = P(\exists n \geq 1 \text{ s.t. } X_n = i \mid X_0 = i) = P(\text{return})$

$f_i = 1 \implies i$ is recurrent (every path leads back to i)

$f_i < 1 \implies i$ is transient

Recurrent State for T_i = time of first return to i:

$\mathbb{E}(T_i \mid X_0 = i) \leq \infty \implies$ positive recurrent

$\mathbb{E}(T_i \mid X_0 = i) = \infty \implies$ null recurrent

Ergodic: an aperiodic, positive recurrent state is ergodic.

A Markov chain is ergodic if all states are ergodic.

Thm (*Existence of Equilibrium Distribution*): for an irreducible, ergodic MC, the limit

$\pi_j = \lim_{n\to\infty} P_{ij}^n$

exists for all j and is independent of i.

1. $\boldsymbol{\pi}$ is the unique solution of $\boldsymbol{\pi} = \boldsymbol{\pi}P$ and $\sum_j \pi_j = 1$
2. Let $N_j(n)$ be the number of visits to state j after n steps. Then $\pi_j = \lim_{n\to\infty} \frac{N_j(n)}{n}$
3. $\pi_j = \frac{1}{m_j}$ where $m_j = \mathbb{E}(T_j \mid X_0 = j)$

Thm (*Time Reversal*): given a MC $(X_n)_{n=0}^N$ with stationary distribution $\boldsymbol{\pi}$ and with $P(X_0 = j) = \pi_j$, let $Y_n = X_{N-n}$. Then $(Y_n)_{n=0}^N$ is a MC with transition probabilities $Q_{ij} = P_{ji}\frac{\pi_j}{\pi_i}$ and stationary distribution $\boldsymbol{\pi}$

Time Reversibility:

$Q_{ij} = P_{ji} \forall i, j \qquad \pi_i P_{ij} = \pi_j P_{ji}$

Entropy: M molecules, B regions $M_1, \ldots, M_B$.

$f_i = \frac{M_i}{M}$

Shannon entropy: $-\sum_{i=1}^B f_i \log f_i$

For an irreducible, ergodic MC, the Shannon entropy increases monotonically iff $\boldsymbol{\pi}$ is uniform.

Relative entropy: $D(\pi_a \parallel \pi_b) = \sum_{i=1}^N \pi_a(i) \log \frac{\pi_a(i)}{\pi_b(i)}$

For an irreducible, ergodic MC, consider two starting distributions π_0 and μ_0. Then $D(\pi_n \parallel \mu_n)$ decreases monotonically.

Random Variables

Distribution	PMF/PDF	Mean	Variance	CF
$\text{Bern}(p)$	$p(1) = p, p(0) = 1-p$	p	$p(1-p)$	$1 - p + pe^{it}$
$\text{Bin}(n,p)$	$p(i) = \binom{n}{i} p^i (1-p)^{n-i}$	np	$np(1-p)$	$(1-p+pe^{it})^n$
$\text{Geom}(p)$	$p(i) = (1-p)^{i-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
$\text{Pois}(\lambda)$	$p(i) = \frac{\lambda^i}{i!}e^{-\lambda}$	λ	λ	$e^{\lambda(e^{it}-1)}$
$\text{Unif}(a,b)$	$f(x) = \frac{1}{b-a} \quad x \in [a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
$\text{Exp}(\lambda)$	$f(x) = \lambda e^{-\lambda x} \quad x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - it}$
$N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{i\mu t - \frac{\sigma^2 t^2}{2}}$

Identities and Approximations

Taylor Expansion of $\cos(x)$:

$\cos(x) \approx 1 - \frac{x^2}{4} + \frac{x^4}{24}$

Stirling's Approximation:

$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

Updated April 27, 2024