MATH 318 Formula Sheet

Probability Theory

Probability Function:

- $0 \leq P \leq 1$
- $P(S) = 1$
- $E_1 \cap E_2 = \emptyset \implies P(E_1 \cup E_2) = P(E_1) + P(E_2)$
- $P(E_1) + P(E_2) = P(E_1 \cup E_2) + P(E_1 \cap E_2)$

Conditional Probability:
$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$
Two events are said to be independent if:
$P(E \cap F) = P(E)P(F)$

Theorem 1. Let $F_1, F_2 \dots F_n$ be a partition of the sample space S. Assume $F_i \cap F_j = \emptyset$ for any $i \neq j$. Then for any event $E \subset S$,

1. $P(E) = \sum_i^n P(E|F_i)P(F_i)$
2. $P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_i^n P(E|F_i)P(F_i)}$ (Bayes' Formula)

Random Variables

Memory-less Property:
$P(X > m + n | X > n) = P(x > m)$

Expectation Value:
$\langle X \rangle = \sum_{i=0}^{\infty} x_i p(X = x_i) = \sum_{i=0}^{\infty} x_i p(x_i)$
$\langle X \rangle = \int_{-\infty}^{\infty} x f(x)\, dx$

Cumulative Distribution Function:
$F(x) = \int_{-\infty}^{x} f(t), dt$

Law of the Unconscious Statistician:
$\langle g(X) \rangle = \int_{-\infty}^{\infty} g(x) f(x)\, dx$

Linearity of Expectation:
$\langle aX + b \rangle = a \langle X \rangle + b$

Moments:
$$n\text{-th moment of } X = \begin{cases} \int_{-\infty}^{\infty} x^n f(x)\, dx \\ \sum_i^{\infty} x_i^n p(x_i) \end{cases}$$
Variance:
$\mathrm{Var}(X) = \left\langle (X - \langle X \rangle)^2 \right\rangle = \left\langle X^2 \right\rangle - \langle X \rangle^2$

Joint Continuity:
$P((X, Y) \in C) = \iint_C f(x, y)\, dx\, dy$

Marginal Distribution:
$P(X \in A) = P(X \in A, Y \in \mathbb{R}) = \int_A \int_{-\infty}^{\infty} f(x, y)\, dy\, dx$

Independence:
If X, Y are independent, then
$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$
$\langle g(X)h(Y) \rangle = \langle g(X) \rangle \langle h(Y) \rangle$

Covariance:
$\mathrm{Cov}(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle$

Correlation Coefficient:
$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1, 1]$$
Cauchy-Swartz Inequality:
$|\langle XY \rangle|^2 \leq \left\langle X^2 \right\rangle \left\langle Y^2 \right\rangle$

Sum of Random Variables:
$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$
$F_{X+Y}(a) = P(X + Y \leq a) =$
$$\int_{-\infty}^{\infty} \int_{-\infty}^{x+y} f_{X+Y}(x, y)\, dx\, dy = \int_{-\infty}^{\infty} f_X(a - y) f_Y(y)\, dy$$
Conditional Probability Distribution:
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$
Conditional Expectation:
$\langle X|Y \rangle = \sum_x x p_{X|Y}(x, y)$
$\langle X|Y \rangle = \int_{\infty}^{\infty} x f_{X|Y}(x, y)\, dx$
$\langle X \rangle = \langle \langle X|Y \rangle \rangle = \sum_y \langle X|Y = y \rangle P(Y = y)$
$\langle X \rangle = \langle \langle X|Y \rangle \rangle = \int_{\infty}^{\infty} \langle X|Y = y \rangle f_Y(y)\, dy$

Characteristic Functions

$\phi_X(t) = \left\langle e^{itX} \right\rangle \quad M(t) = \left\langle e^{tX} \right\rangle$

Extracting Moments: $\frac{d^n}{dx^n}|_{t=0}\phi(t) = \langle i^n X^n \rangle$

Inversion Theorem:
$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t) e^{-itx}\, dt$

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

Convergence of Random Variables

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

Thm (*Continuity Theorem*):
Let $X_1, X_2, \dots$ be random variables with CDFs $F_1, F_2, \dots$ and characteristics functions $\phi_1, \phi_2, \dots$. Then

· If $F_n \to F$, where F is the CDF of some random variable X, then $\lim_{n\to\infty} \phi_n(t) = \phi(t)$.

· If $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ and $\phi(t)$ is continuous at $t = 0$, then ϕ is the characteristic function of some random variable X and $F_n \to X$ and $X_n \xrightarrow{D} X$.

Thm (*Weak Law of Large Numbers*): Let $X_1, X_2, \dots X_n$ be iid random variables. Assume $\langle X \rangle = \mu < \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then $\frac{S}{n} \xrightarrow{D} \mu$.

Thm (*Central Limit Theorem*): Let X_i be iid random variables. Let $\langle X_i \rangle = \mu < \infty$, $\mathrm{Var}(X_i) = \sigma^2 < \infty$, and $S_n = \sum_{i=1}^n X_i$. Then
$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0, 1).$$

Statistical Estimation

Def (*Estimator*): A function of the data that is used to estimate the unknown parameter.

Estimator of the mean μ:
$$\text{Sample Mean } \bar{X} = \frac{1}{n} \sum_1^n X_i$$
Estimator of the variance σ^2:
$$\text{Sample Variance } S^2 = \frac{1}{n - 1} \sum_{i=1}^n (X_i - \bar{X})^2$$
Def (*Confidence Interval*): For some interval $A \subset \mathbb{R}$, an estimator is in the a% confidence level if $P(\bar{X} \in A) = a$%.

Random Walks

Symmetric random walks in $\mathbb{Z}^d$:
Number of visits to the origin M: $\langle M \rangle = (1 - u)^{-1}$.
Probability walk returns to the origin $u = 1 - 1/\langle M \rangle$.
If $u = 1$, the walk is recurrent, otherwise transient.

Markov Chains

Transition Matrix P: Rows add to 1.

One-Step Transition Probability:
$P_{ij} = P(X_{n+1} = j | X_n = i) \quad \sum_j P_{ij} = 1$

N-Step Transition Probability:
$P^n = P(X_{l+n} = j | X_l = i)$

Chapman-Kolmogorov Equation:
$P_{ij}^{n+m} = \sum_k P_{ik}^n P_{kj}^n$
$P^{n+m} = P^n P^m$

Classification of States:

- A 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.
- If i is recurrent and j is accessible from i, $i \leftrightarrow j$.
- If i is recurrent and $i \leftrightarrow j$, then j is also recurrent.

Irreducibility:
A Markov chain is irreducible if there is only one state (all states communicate).
Periodicity of state i:
Period $d = \gcd\{n \geq 1 : P_{ii}^n > 0\}$
$d = 1 \implies i$ is aperiodic
Transience and Recurrence:
$f_i = P(\exists n \text{ s.t. } X_n = i | X_0 = i)$
$f_i = 1 \implies i$ is recurrent (every path leads to i)
$f_i < 1 \implies i$ is transient
Recurrent State for T_i = time of first return to i:
$\langle T_i | X_0 = i \rangle < \infty \implies$ positive recurrent
$\langle T_i | X_0 = i \rangle = \infty \implies$ null recurrent
Def (*Ergodic*): A aperiodic, positive recurrent state is called ergodic. A Markov chain is ergodic if all its states are ergodic.
Thm (*Existence of Equilibrium Distribution*): For an irreducible, ergodic Markov chain, the limit
$\pi_j = \lim_{n\to\infty} P_{ij}^n$
exists for all j and is independent of state 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. $\pi_j = 1/m_j$ where $m_j = \langle T_j | X_0 = j \rangle$

Time Reversal:
Given a Markov chain $(X_n)_{n=0}^N$ 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 Markov chain with transition probabilities $Q_{ij} = P_{ji}\frac{\pi_j}{\pi_i}$ and stationary distribution $\boldsymbol{\pi}$.
Def (*Time Reversibility*): A markov chain is time reversible if $Q_{ij} = P_{ij}\ \forall i, j$. In this case, $\pi_i P_{ij} = \pi_j P_{ji}$.

Updated April 15, 2020
https://github.com/DonneyF/formula-sheets

Random Variables

Distribution	Mass/Density Function	Mean	Variance	Characteristic Function
Binomial(n,p)	$p(i) = \binom{n}{i} p^i (1-p)^{n-i}$	np	$np(1-p)$	$(1-p+e^{it})^n$
Geometric(p)	$p(k) = (1-p)^{k-1}p$	$1/p$	$\frac{1-p}{p^2}$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
Poisson(λ)	$p(i) = \frac{\lambda^i}{i!}e^{-\lambda}$	λ	λ	$e^{\lambda(e^{it}-1)}$
Uniform(a,b)	$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{ita}-e^{itb}}{it(b-a)}$
Exponential(λ)	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$1/\lambda$	$1/\lambda^2$	$\frac{\lambda}{\lambda - it}$
Normal(μ,σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{i\mu t - \sigma^2 t^2/2}$

Reserved

Reserved