PwA Cheatsheet

Common Distributions

Discrete				
Name	**pmf**	**cdf**	**mean**	**variance**
Binomial(n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$	$F(k;n,p)=\Pr(X\le k)=\sum_{i=0}^{\lfloor k\rfloor}\binom{n}{i}p^i(1-p)^{n-i}$	np	$np(1-p)$
Neg. Binomial(r,p)	$\binom{i-1}{r-1}p^r(1-p)^{i-r}$	-	$\frac{r}{p}$	$r\frac{1-p}{p^2}$
Bernoulli(p)	$\begin{cases} q=(1-p) & \text{for } k=0 \\ p & \text{for } k=1 \end{cases}$	$\begin{cases} 0 & \text{for } k<0 \\ 1-p & \text{for } 0\le k<1 \\ 1 & \text{for } k\ge 1 \end{cases}$	p	$p(1-p)$
Uniform(a,b)	$\frac{1}{n}, n=b-a+1$	$\frac{\lfloor k\rfloor-a+1}{n}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Geometric(p)	$p(1-p)^{i-1}$	$1-(1-p)^i$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Hypergeometric(N,K,n) "k successes ⊂ N, K suc ∈ N"	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	-	$n\frac{K}{N}$	$n\frac{K}{N}\frac{(N-K)}{N}\frac{N-n}{N-1}$
Poisson(λ)	$\frac{\lambda^k e^{-\lambda}}{k!}$	$e^{-\lambda}\sum_{i=0}^{\lfloor k\rfloor}\frac{\lambda^i}{i!}$	λ	λ

Continuous				
Name	**pdf**	**cdf**	**mean**	**variance**
Uniform(a,b)	$\begin{cases} \frac{1}{b-a} & \text{for } x\in[a,b] \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 0 & \text{for } x<a \\ \frac{x-a}{b-a} & \text{for } x\in[a,b) \\ 1 & \text{for } x\ge b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal(μ,ω^2)	$\frac{1}{\sqrt{2\sigma^2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\frac{1}{2}\left[1+\text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$	μ	σ^2
Exponential(λ)	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$
Hazard/Failure Rate Functions				
Survival	**Hazard**	**Distribution**	**Rate**	**Book**
$\bar{F}(t)=1-F(t)$	$\lambda(t)=\frac{f(t)}{\bar{F}(t)}$	$F(t)=1-exp\{-\int_0^t\lambda(t)dt\}$	λ	p217

Events

Sample Space	$S=\{all\ possible\ outcomes\}$
Event	$E\subset S$
Union (either or both)	$E\cup F$
Intersection (both)	$E\cap F$ or EF
Complement	$E^C=S\backslash E\Rightarrow P(E^C)=1-P(E)$
Inclusion-Exclusion	$\hookrightarrow P(A\cup B)=P(A)+P(B)-P(A\cap B)$
DeMorgan's Law	1. $(E_1\cup\ldots\cup E_n)^C=E_1^C\cap\ldots\cap E_n^C$ 2. $(E_1\cap\ldots\cap E_n)^C=E_1^C\cup\ldots\cup E_n^C$
Axioms	1. $0\le P(E)\le 1$ 2. $P(S)1$ 3. For mutually excl. events $A_i, i\ge 1$: $P(\cup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$
Finite S, Equal Probability for all point sets:	$P(A)=\mid A\mid\div\mid S\mid$
Odds of Event	$\alpha=\frac{P(A)}{P(A^C)}=\frac{P(A)}{1-P(A)}$

Conditional Probability and Independence I

Conditional Probability	$P(F\mid E)=\frac{P(F\cap E)}{P(E)}$
Independence if	$P(F\cap E)=P(F)P(E)$
Multiplication Rule	$P(E_1E_2\cdots E_n)=P(E_1)P(E_2\mid E_1)\cdots P(E_n\mid E_1\cdots E_{n-1})$
Bayes Formula (simple)	$P(A\mid B)=\frac{P(B\mid A)P(A)}{P(B)}$.
Bayes Formula (full)	$P(A_i\mid B)=\frac{P(B\mid A_i)P(A_i)}{\sum_j P(B\mid A_j)P(A_j)}$.
Conditional pmf (discrete)	$p_{X\mid Y}(x\mid y)=\frac{p(x,y)}{p_Y(y)}$
Conditional pdf (discrete)	$F_{X\mid Y}(x\mid y)=\sum_{a\le x}p_{X\mid Y}(a\mid y)$
Conditional Density (continuous)	$f_{X\mid Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}$
Conditional Probabilities (continuous)	$P\{X\in A\mid Y=y\}=\int_A f_{X\mid Y}(x\mid y)dx$

Random Variables (Discrete)

Distribution Function	$F(x)=P\{X\le x\}$
Probability Mass Function	$p(x)=PX=x$
Joint Probability Mass Function	$\text{P}(X=x \text{ and } Y=y)$ $=\text{P}(Y=y\mid X=x)\cdot\text{P}(X=x)$ $=\text{P}(X=x\mid Y=y)\cdot\text{P}(Y=y)$
Expectation	$E[X]=\sum_{x:p(x)>0}xp(x)$
$\hookrightarrow note:$	$E[g(X)]=\sum_{x:p(x)>0}g(x)p(x)$
Variance	$Var(X)=E[(X-E[X])^2]$ $=E[X^2]-(E[X])^2$
Standard Derivation	$\sigma=\sqrt{Var(X)}$
Covariance	$Cov(X,Y)=E[(X-E[X])(Y-E[Y])]$ $=E[XY]-E[X]E[Y]$
Moment Gener. Function	$M(t)=E[e^{tX}]$ (same for continuous RVs)

Random Variables (Continuous) I

Probability Density Function	f such that $P\{X\in B\}=\int_B f(x)dx$
Distribution Function	F such that $\frac{d}{dx}F(x)=f(x)$
Expectation	$E[X]=\int_{-\infty}^{\infty}xf(x)dx$
$\hookrightarrow note:$	$E[g(X)]=\int_{-\infty}^{\infty}g(x)f(x)dx$
Variance	$Var(X)=E[(X-E[X])^2]$ $=E[X^2]-(E[X])^2$
Standard Derivation	$\sigma=\sqrt{Var(X)}$
Covariance	$Cov(X,Y)=E[(X-E[X])(Y-E[Y])]$ $=E[XY]-E[X]E[Y]$
Joint Probability Mass Function	$P\{(X,Y)\in C\}=\iint_{(x,y)\in C}f(x,y)dxdy$ $P\{X\in A, Y\in B\}=\int_B\int_A f(x,y)dxdy$

Random Variables (Continuous) II

Marginal pmfs	$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy$ $f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx$

More on Expectation, Variance, ..

$E[X+Y]=E[X]+E[Y]$

$E[\alpha X]=\alpha E[X]$

$Var(X+a)=Var(X)$

$Var(aX+b)=a^2Var(X)$

$$\begin{aligned}Var(X+Y)&=E[(X+Y)^2]-(E[X+Y])^2\\&=E[X^2+2XY+Y^2]-(E[X]+E[Y])^2\\&=E[X^2]+2E[XY]+E[Y^2]-\\&\quad(E[X])^2-2E[X]E[Y]-(E[Y])^2\\&=Var(X)+Var(Y)+2(E[XY]-E[x]E[Y])\\&=Var(X)+Var(Y)+2(Cov(X,Y))\end{aligned}$$

Independence
$\Rightarrow E[f(X)g(Y)]=E[f(X)]E[g(Y)]$
$\Rightarrow E[XY]=E[X]E[Y]$
$\Rightarrow Cov(X,Y)=0$
$\Rightarrow Var(X+Y)=Var(X)+Var(Y)$

Correlation $corr(X,Y)=\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

1. $-1\le\rho(X,Y)\le 1$
2. Independence $\Rightarrow\rho(X,Y)=0$
   $Y=mX+cm, m\neq 0\ and\ c:$
3. $m>0\Rightarrow\rho(X,Y)=1$
   $m<0\Rightarrow\rho(X,Y)=-1$

$E[X]=E[E[X\mid Y]]$

Disc.: $E[X]=\sum_y E[X\mid Y=y]P\{Y=y\}$

Cont.: $E[X]=\int_{-\infty}^{\infty}E[X\mid Y=y]f_y(y)dy$

Combinatorial Analysis

Order matters and k = n	Permutation
Order does matter and k < n	Variation
Order does not matter and k < n	Combination

Counting

Basic Counting Principle	Experiments $E_1, E_2, ..E_r$ with $n_1, n_2, ..n_r$ possible outcomes. Total outcomes: $\prod_i^r n_i$
Permutations (without Repeats)	$n! = n \cdot (n-1) \cdot \ldots \cdot 1$
Permutations (with Repeats)	$\frac{n!}{k!} = n \cdot (n-1) \cdot \ldots \cdot (k+1)$
Variations (without Repeats)	$n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}$
Variations (with Repeats)	$\underbrace{n \cdot \ldots \cdot n}_{k\text{-times}} = n^k$
Combinations (without Repeats) "Binomial Coefficient"	$\frac{n!}{(n-k)!\,k!} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} = \binom{n}{n-k} = \binom{n}{k}$
Multinominal Coefficient	$\frac{n!}{n_1! n_2! \ldots n_r!} = \binom{n!}{n_1, n_2, \ldots, n_r}$ "divide n into r non-overlapping subgroups of sizes n1,n2,.."
Combinations (with Repeats)	$\frac{(n+k-1)!}{(n-1)!\,k!} = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}$

Limit Theorems

Central Limit Theorem	$Z_n = \frac{((X_1 + X_2 + \ldots + X_n) - n\mu)}{\sigma\sqrt{n}}$ Then as $n \to \infty$ $P(Z_n \leq x) \to \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}u^2) du$ $i.e. P(Z_n \leq x) \to P(Y \leq X)$ where $Y \sim N(0,1)$
Weak Law of Large Numbers	$E[X_i] = \mu \quad Var(X_i) = \sigma^2$ $s_n = \frac{1}{n}(X_1 + \ldots + X_n)$ then for any $\epsilon > 0$ $\lim_{n\to\infty} P(\mid s_n - \mu \mid > \epsilon) = 0$
Strong Law of Large Numbers	$P\{lim_{n\to\infty}(X_1 + X_2 + \ldots + X_n) \div n = \mu\} = 1$
Markov's Inequality	$p\{x \geq a\} \leq \frac{E[X]}{a}$
Chebyshev's Inequality	$E[Y^2] < \infty, \forall a > 0.$ $P(\mid Y \mid \geq \frac{1}{a^2} E[Y^2])$
$\hookrightarrow$	$P\{\mid X - \mu \mid \geq a\} \leq \frac{\sigma^2}{a^2}$
One-sided Chebyshev (mean 0)	$P\{X \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$
Chernoff Bounds	$P\{X \geq a\} \leq e^{-ta} M(t) \quad t > 0$ $P\{X \leq a\} \leq e^{-ta} M(t) \quad t < 0$

Markov Chains

Discrete

$P_{i,j} = P(\text{system is in state } j \text{ at time } n+1 \mid \text{system in state } i \text{ at time } n)$

$$\text{Transition Matrix: } P = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{pmatrix}$$

Probability vector $\pi^{(n)}$: Probabilities that we are in state i at n.

$$\pi^{(n+1)} = \pi^{(n)} P$$

$$\pi^{(n)} = \pi^{(0)} P^n$$

A Markov Chain is ergodic(aperiodic and irreducible) iff there exists $n \in \mathbb{N}^+$ such that P^n has no zero entries. It then has a Steady State Probability vector $\pi = lim_{n\to\infty} \pi^{(n)}$ independent of $\pi^{(0)}$.

1. $\pi_0 + \pi_1 + \ldots + \pi_N - 1 = 1$
2. $\pi = \pi P$

Continuous

Poisson

$$P(\tilde{N}(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \text{ if}$$

1. For any fixed t, $\tilde{N}(t)$ is a discrete RV
2. $\tilde{N}(0) = 0$
3. # of events in disjoint intervals are independent
4. $\tilde{N}(t+h) - \tilde{N}(t) = \#$ of events in $[t, t+h]$ for $h \to 0$
5. $P(\tilde{N}(h) = 1) = P(\text{event occurs in } [t, t+h]) = \lambda h + E(h)$ $(E(h)/h \to 0$, as $h \to 0)$
6. $\frac{1}{h} P(\tilde{N}(h) \geq 2) \to 0$, as $h \to 0$

Birth-Death

Birth Rates $\lambda_{i,i+1} = b_i$
Death Rates $\lambda_{i,i-1} = d_i$
$\lambda_{i,j} = 0$, otherwise

$\to$ Have steady state prob. vector if b_is and d_is are non-zero and we have a finite number of states.

1. $\pi_0 + \pi_1 + \ldots + \pi_N - 1 = 1$
2. $\pi_j = \frac{b_0 \cdots b_{j-1}}{d_0 \cdots d_{j-1}} \pi_0$

M/M/S Queue

Customers arrive with Poisson Process rate λ, S servers, Service time exponentially distributed with mean $\frac{1}{\mu}$. State j = j customers in queue,

$$b_j = \lambda, \ d_j = \begin{cases} j\mu, & j = 1, 2, \ldots, S \\ S\mu, & j \geq S \end{cases}$$

$\to$ Has steady state prob. vector if $\lambda < S\mu$.

M/M/1 Queue

$\pi_j = (1 - \frac{\lambda}{\mu})(\frac{\lambda}{\mu})^j$, Mean Queue Length $E[J] = \frac{\lambda}{\mu - \lambda}$

Surprise, Uncertainty & Entropy

Entropy	$H(X) := -\sum_k p_x(x_k) log_2 p_x(x_k)$ $(0 log_2(0) := 0)$
Surprise	$S(X = x_k) = -log_2 p_x(x_k)$
$\hookrightarrow$ Properties	1. $S(1) = 0 \neq S(0)$ (which is undefined) 2. S decreases: $p < q \Rightarrow S(q) < S(p)$ 3. $S(pq) = S(p) + S(q)$ If S is continuous and these are satisfied, $\exists \mathscr{C} > 0. \forall p \in [0,1], S(p) = -\mathscr{C} log(p)$
Average Uncertainty	$H(X,Y) := -\sum_j \sum_k p_{X,Y}(x_j, y_k) log_2 p_{X,Y}(x_j, y_k)$
Uncertainty of X given Y	$H_{Y=y_k}(X) := -\sum_j p_{X\mid(Y=y_k)}(x_j) log_2 p_{X\mid(Y=y_k)}(x_j)$
Conditional Entropy	$H_Y(X) := \sum_k H_{Y=y_k}(X) p_Y(y_k)$

Coding Theory

Code $\mathfrak{C}$	A map from $\{x_k\} \subset \mathbb{R}$ into sequences of 0's and 1's. Sequences are called code words.
Code Word length	$x_k \mapsto 0111 \quad \Rightarrow n + k =$
Expected Length of Code $\mathfrak{C}$	$E[\mathfrak{C}] = \sum_k n_k p_k = \sum_k n_k P(X = x_k)$
Acceptable Code	No code word extends another one: ✗: $x_1 \mapsto 0$, $x_2 \mapsto 00$; ✓: $x_1 \mapsto 0$, $x_2 \mapsto 10$
Noiseless Coding Theorem	For any acceptable code assigning n_k bits to x_k the following holds: $E[\mathfrak{C}] = \sum_k n_k p_k \geq H(X) = -\sum_k p_k log_2 p_k$ Where $p_k = P(X = x_k)$ and n_k length of a codeword associated with x_k
$\hookrightarrow$ Thrm	For any discrete RV X there exists an acceptable code with the expected length $E[\mathfrak{C}] = L$ such that $H(X) \leq L < H(X) + 1$

Algorithm for finding an acceptable code with expected length $H(X) \leq E[\mathfrak{C}] = L < H(X) + 1$ for discrete RV X:

1. Let n_j be the integer satisfying $-log_2 p_j \geq n_j < -log_2 p_j + 1$
2. Find any acceptable code assigning n_j bits to x_j

There is no unique nearly-optimal code in general. Optimal or nearly-optimal coding depends on the pmg of X.

Common Moment Generating Functions M(t)

Binomial	$(pe^t + 1 - p)^n$
Neg. Binomial	$[(pe^t) \div (1 - (1-p)e^t)]^r$
Poisson	$exp(\lambda(e^t - 1))$
Uniform	$(e^{tb} - e^{ta}) \div (t(b-a))$
Exponential	$\lambda \div (\lambda - t)$
Normal	$exp(\mu t + ((\sigma^2 t^2) \div 2))$