ST2131 CheatSheet

by Zachary Chua

Counting

Sample Space: Set of all possible outcomes of an experiment

Event: Subset of sample space

Naive definition of probability: Assumes all outcomes are equally likely

$P(A) = \frac{|A|}{|S|}$, where A is an event

Limitations — not equally likely, infinitely many outcomes

Note: if numerator ordered, then denom also ordered and vice versa

Product Rule: compound experiment with sub-experiments A and B, A has a possible outcomes B has b possible outcomes, compound experiment has ab possible outcomes

Binomial Coefficient: $nCk = \frac{n!}{(n-k)!k!}, 0 \leq k \leq n$

Sampling

1. **Order matters, w replacement**: n^k
2. **Order matters, w/o replacement**: $n(n-1)\dots(n-k+1)$
3. **Order doesn't matter, w replacment**: $\binom{n+k-1}{k}$
   - Stars and Bars, n boxes, k indistinguishable balls
   - no. of nonneg integer solns to $x_1 + x_2 + \cdots + x_n = k$
4. **Order doesn't matter, w/o replacement**: $\binom{n}{k}$

Vandermonde's Identity: $\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}$

- m parrots, n eagles, select k birds and select j parrots, select k - j eagles

Axioms of Probability

1. $P() = 0, P(S) = 1$
2. $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ if A_n are disjoint events
   - Disjoint = mutually exclusive and non-overlapping

Inclusion-Exclusion: prob of union

$P(\bigcup_{i=1}^{n} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n)$

Try to look for symmetry to remove terms, or intersections = 0, then higher intersections = 0 also

Probability of intersection

$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)\dots P(A_n|A_1, \dots, A_{n-1})$

Bayes: $P(A|B) = \frac{P(B|A)(A)}{P(B)}$

Note: $P(A|B) \neq P(B|A)$, Prosecutor's fallacy

Law of Total Probability (LOTP): $P(B) = \sum_{j=1}^{n} P(B|A_j)P(A_j)$

if $A_1 \dots A_n$ is a partition of S

Conditional Independence: $P(A, B|C) = P(A|C)P(B|C)$

Note: Indep $\not\rightarrow$ Cond Indep and vice versa

Difference Equation: $X_n = aX_{n-1}bX_{n-2}$

Guess $X_n = \alpha^n, \alpha \neq 0$

then $\alpha^2 - \alpha a - b = 0$, let α_1, α_2 be the roots

then α_1^n, α_2^n, general solution is $c_1\alpha_1^n + c_2\alpha_2^n$

Discrete Distributions

Random Variable: function that maps for sample space to real line

Every R.V. has a distribution, whic specifies all probabilities for that r.v

Note: PMF ≥ 0, sums to 1

Expectation: $E(X) = \sum_x xP(X = x)$

1. $E(cX) = cE(X)$
2. $E(X + Y) = E(X) + E(Y)$

Indicator RV: I(A) = 1 if A, 0 otherwise, $E(I(A)) = P(A)$

LOTUS: $Y = g(X), E(Y) = \sum_x g(x)P(X = x)$

Variance: Distance of r.v from its mean

1. $Var(X) = E((X - \mu)^2) = E(X^2) - EX^2$
2. $Var(X) \geq 0$, if equal to 0, then r.v is constant
3. $Var(cX) = c^2 Var(X)$
4. $Var(X + Y) \neq Var(X) + Var(Y)$, unless X and Y are independent
5. $Var(X + c) = Var(X)$

Bernoulli: $X \sim Bern(p)$

PMF: $P(X = 1) = p, P(X = 0) = 1 - p, 0 \leq p \leq 1$

Expectation: p

Binomial: $X = \#$ of success, $X \sim Bin(n, p)$

PMF: $\binom{n}{k} p^k q^{n-k}, k \in 0, \dots, n$, 0 otherwise

Expectation: np (sum of bernoullis)

Hypergeometric: $X \sim HGeom(w, b, n)$, number of white balls in n balls

PMF: $P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$

Expectation: $n(\frac{w}{w+b})$, sum of dependent Bernoullis

Variance: $\frac{w+b-n}{w+b-1}\mu(1 - \frac{\mu}{n})$, Covariance with itself

Geometric: X = # failures before first success (excl first success)

indep trials, each with prob p of success

$X \sim Geom(p)$

PMF: $P(X = k) = p(1-p)^k$

Expectation: q/p

Variance: q/p^2

First Success: Geometric + 1

PMF: $P(X = k) = p(1-p)^{k-1}$

Expectation: $\frac{1}{p}$

Variance: q/p^2

Negative Binom: general geom

PMF: $P(X = k) = \binom{k+r-1}{r-1} p^r (1-p)^k$

Expectation $= \frac{rq}{p}$

Variance $= \frac{rq}{p^2}$, can sum indiv geom because indep

Poisson: when counting rare things w no predetermined upper bound

$X \sim Pois(\lambda)$, wher λ is a +ve real number

Support: $\{0, 1, 2, 3, \dots\}$

PMF: $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$

Expectation: λ

Variance: λ

Note: sum of indep poissons is poisson, $X_1 \sim Pois(\lambda_1), X_2 \sim Pois(\lambda_2)$ then $X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)$

Possion Approx

Events $A_1, A_2, \dots, A_n$, n LARGE (indep or *slight* dependencies)

let $P(A_J) = p_j$ (small, < 0.01)

X = $\#A_j$ that occur, then X is approx poisson, $\lambda = p_1 + p_2 + \cdots + p_n$

Continuous Random Variables

Support is real line or subinterval

CDF is differentiable, PDF is derivative. PDF $\neq P(X = x) = 0$, uncountably many x

Integrate PDF to get probability.

Expectation: $\int_{-\infty}^{\infty} xf(x)dx$

LOTUS: $\int_{-\infty}^{\infty} g(x)f(x)dx$

Standard Normal: $N(0, 1)$

PDF: $\frac{1}{2\pi}e^{-x^2/2}$

CDF: $\int_{-\infty}^{x} f(t)dt = \Phi(x)$

Symmetric: $\Phi(-z) = 1 - \Phi(z)$

General Normal: $N(\mu, \sigma^2)$

Transformation from Z: let $X = \mu + \sigma Z$, then $X \sim N(\mu, \sigma^2)$

Transformation to Z: $Z = \frac{X-\mu}{\sigma}$, then $Z \sim N(0, 1)$

CDF: $\Phi(\frac{x-\mu}{\sigma})$

PDF: $\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma}), \phi(z) = \frac{1}{2\pi}e^{-z^2/2}$

Uniform Distribution: $Unif(a, b)$

1D: probability $\propto$ length

PDF: $cifa < x < b, 0otherwise$, $c = \frac{1}{b-a}$

CDF: $\frac{x-a}{b-a}$, 0, 1

Log-Normal Distribution: $LogNormal(\mu, \sigma^2)$, Log IS normal

If have product of +ve r.v., take log, changes to sum.

$X \sim N(\mu, \sigma^2)$, $Y = e^x$, $log(Y) = X$

then $Y \sim LogNormal(\mu.\sigma^2)$

CDF: $\Phi(\frac{\log(y)-\mu}{\sigma})$

PDF: $\phi(\frac{\log(y)-\mu}{\sigma})\frac{1}{\sigma}\frac{1}{y}$

Exponential Distribution: $Expo(\lambda)$

CDF: $1 - e^{-\lambda x}$, $x > 0$

PDF: $\lambda e^{-\lambda x}$, $x > 0$

Memoryless: $P(x > s + t | x > s) = P(x > t)$

If $X \sim Expo(\lambda)$, then $Y = \lambda X \sim Expo(1)$

Expectation: $\frac{1}{\lambda}$

Variance: $\frac{1}{\lambda^2}$

Gamma Distribution:

Gamma Function $\Gamma(a) = \int_0 \infty x^a e^{-x} \frac{dx}{x} = (a-1)!$, for $a > 0$

Gamma(a, 1) PDF: $\frac{1}{\Gamma(a)}x^a e^{-x}\frac{1}{x}, x > 0$

$X \sim Gamma(a, 1)$, $Y = \frac{1}{\lambda}X \sim Gamma(a, \lambda)$

$f_Y(y) = \frac{1}{\Gamma(a)}(\lambda y)^a e^{-\lambda y}\frac{1}{y}$

Gamma(1, 1) = Expo(1)

Possion Process Interpretation: If $X_1, \dots, X_n$ iid $Expo(\lambda)$ then $X_1 + \cdots + X_n \sim Gamma(n, \lambda)$

Beta Distribution: $Beta(a, b)$

PDF: $f(x) = cx^{a-1}(a-x)^{b-1}$, $0 < x < 1, a > 0, b > 0$

$c = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$

Unif(0,1) = Beta(1, 1)

Bank Post Office: $X \sim Gamma(a, \lambda), Y \sim Gamma(b, \lambda)$, independent

Results:

1. $X + Y$ indep of $\frac{X}{X+Y}$
2. $X + Y \sim Gamma(a + b, \lambda)$
3. $\frac{X}{X+Y} \sim Beta(a, b)$
4. Beta normalising constant

Universality of the Uniform

1. Let $U \sim Unif(0, 1)$, let F be a CDF which is continuous and strictly $\uparrow$
   1.1 Then $F^{-1}(U) \sim F$
2. let $X \sim F$, let $U = F(X)$, then $U \sim Unif(0, 1)$

Poisson Process

$N_t = \#$ of arrivals in [0, t]

1. $N_t \sim Pois(\lambda t)$
2. # of arrivals in disjoin intervals are indep

Intervals between arrivals are iid $Expo(\lambda)$

Joint Distributions

Covariance and Correlation

Covariance $Cov(X, Y)$: $E((X - EX)(Y - EY)) = E(XY) - E(X)(Y)$

Note: $Cov(X, X) = Var(X)$

Correlation ($Corr(X,Y)$): Covariance between standardised X and Y, $\frac{Cov(X,Y)}{SD(X)SD(Y)}$

Note: Independent implies Uncorrelated (but converse not true), Cov = 0 if indep

Properties:

1. $Cov(X, Y + c) = Cov(X, Y)$
2. $Cov(aX, Y) = aCov(X, Y)$
3. $Cov(X, Y) = Cov(Y, X)$
4. $Cov(X, X) = Var(X)$
5. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$ (like distributive law)
6. Bilinearity: $Cov(X+Y, Z+W) = Cov(X, Z)+Cov(X, W)+Cov(Y, Z)+Cov(Y, W)$

	Two discrete r.v.s	Two continuous r.v.s
Joint CDF	$F_{X,Y}(x,y) = P(X \le x, Y \le y)$	$F_{X,Y}(x,y) = P(X \le x, Y \le y)$
Joint PMF/PDF	$P(X = x, Y = y)$ • Joint PMF is nonnegative. • Joint PMF sums to 1. • $P((X,Y) \in A) = \sum\sum_{(x,y)\in A} P(X = x, Y = y)$.	$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$ • Joint PDF is nonnegative. • Joint PDF integrates to 1. • $P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)dxdy$.
Marginal PMF/PDF	$P(X = x) = \sum_y P(X = x, Y = y)$ $= \sum_y P(X = x \mid Y = y)P(Y = y)$ LOTP (disjoint cases)	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$ $= \int_{-\infty}^{\infty} f_{X\mid Y}(x\mid y)f_Y(y)dy$ continuous LOTP.
Conditional PMF/PDF	$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$ $= \frac{P(X = x \mid Y = y)P(Y = y)}{P(X = x)}$ Bayes Rule	$f_{Y\mid X}(y\mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ $= \frac{f_{X\mid Y}(x\mid y)f_Y(y)}{f_X(x)}$
Independence knowing value of x does not give any info about y	$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$ CDFs $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all x and y. PMFs	$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$ PDFs $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x and y.

Note: Careful with limits of integration

Multinomial:

n indiv, each in exactly 1 category, p_i is prob of belonging to category i

sum of p_is = 1

$X_i = \#$ of individuals in category i

$\vec{X} \sim Mult_k(n, \vec{p})$, $X_j \sim Binom(n, p_j)$, dependent binomials

Lumping Property: combine categories, still multinomial, probs add together

Moments

Definition: nth moment of X is $E(X^n)$

1st moment: $E(X)$

2nd moment: $E(X^2)$, Variance if $E(X) = 0$

3rd moment: $E(X^3)$, related to skewness

4th moment: $E(X^4)$, related to kurtosis

Moment Generating Function (MGF):

MGF of X is M, $M(t) = E(e^{tX})$

1. Used to calc. Moments
2. Determines a distribution uniquely
3. Works well w sum of indep r.v, eg. X, Y indep
   - Then $M_{X+Y}(t) = M_X(t)M_Y(t)$

Calculating Moments

$M(t) = E(e^{tX}) = E(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}) = \sum \frac{E(X^n)t^n}{n!}$

- nth moment is the coef. of $\frac{t^n}{n!}$ in taylor expansion of $M(t)$
- $E(X^n)$ is the nth derivative of M evaluated at 0, $E(X^n) = M^{(n)}(0)$

Note: $M(0) = 1$

Transformations and Convolutions

Transformations

1. **Case 1: 1 dimension**

$Y = g(X)$, where g is differentiable, strictly increasing, X is continuous

- Then, $f_Y(y) = f_X(x)|\frac{dx}{dy}|$, as a function of Y

Proof:

1. $P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(x)$
2. Differentiate both sides wrt to y, $f_Y(y) = f_X(x)|\frac{dx}{dy}|$

2. **Case 2: n dimensions**

$\vec{Y} = g(\vec{X})$, $\vec{X} = (X_1, X_2, \ldots, X_n)$

- Then, $f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(\vec{x})|\frac{\partial \vec{x}}{\partial \vec{y}}|$
- Absolute value of determinant of the jacobian matrix

eg. $\frac{d(u,v)}{d(x,y)} = \begin{pmatrix} \frac{du}{dx} & \frac{du}{dy} \\ \frac{dv}{dx} & \frac{dv}{dy} \end{pmatrix}$

Convolutions: $X + Y = T$, X and Y indep

1. Story, eg. with binomials
2. MGF, but might not exist, hard to convert to PDF
3. Convolution Sum or Integral

1. **Discrete Case**: LOTP

$$P(T = t) = \sum_x P(Y = t - x)P(X = x)$$

2. **Continuous**: Convolution Integral

$$f_T(t) = \int_{-\infty} \infty f_Y(t - x)f_X(x)dx$$

Order Statistics

PDF: $f_{X_{(j)}}(x) = n\binom{n-1}{j-1}f(x)F(x)^{j-1}(1 - F(X))^{n-j}$

Conditional Expectation

Conditional Expectation given an Event A

1. **Discrete**: $E(Y|A) = \sum_y yP(Y = y|A)$
2. **Continuous**: $E(Y|A) = \int_{-\infty}^{\infty} yf(y|A)dy$

Note: Linearity and other rules still hold

Law of Total Expection: LOTE

$$E(Y) = \sum_{j=1}^{n} E(Y|B_j)P(B_j)$$

Example: Waiting for HH vs HT

HT:

1. wait for first H: FS(1/2)
2. wait for first T after H: FS(1/2)

$E(HT) = E(FS(1/2)) + E(FS(1/2)) = 2 + 2 = 4$

HH: cannot "build up progress" like HT

$E(HH) = E(HH|H_1)(1/2) + E(HH|T_1)(1/2)$

$= (1/2(2) + 1/2(c + 2))(1/2) + (c + 1)(1/2)$, where c is E(HH)

$= 6$

Conditional Expectation given an r.v: $g(X) = E(Y|X)$

Properties:

1. If indep, $E(Y|X) = E(Y)$
2. $E(h(X)|X) = h(X)$ (completely dependent)
3. $E(h(X)Y|X) = h(X)E(Y|X)$ (taking out whats known)
4. Linearity

Adam's Law (LOTE): $E(Y) = E(E(Y|X))$

Eve's Law (total var): $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$

Inequalities

1. **Markov**: $P(|X| \ge a) \le \frac{E|X|}{a}$, $a > 0$
   1.1 Proof using indicator r.v.
2. **Chebyshev**: $P(|X - \mu| \ge c\sigma) \le \frac{1}{c^2}$
3. **Cauchy-Schwarz**: $E|XY| \le \sqrt{E(X^2)E(Y^2)}$
   3.1 find exact use 2D LOTUS
   3.2 find distribution use Jacobians transformation from $(x, y) \to (xy, x)$
4. **Jensen**: If g is convex: $E(g(X)) \ge g(E(X))$
   4.1 Convex: $g''(x) \ge 0$ or
   4.2 take any two points on function, line through them is above curve
   4.3 Variance > 0 follows from this
   4.4 if concave: $E(h(X)) \le h(EX)$

Limit theorems and Law of large numbers

Sample Mean:

$X_1, X_2, \ldots$ iid, mean μ, variance σ^2

$\overline{X_n} = \frac{X_1+\cdots+X_n}{n}$, note sample mean is a r.v

$E(\overline{X_n}) = \mu$, $Var = \sigma^2/n$

Sample Variance:

$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X_n})^2$

$E(s^2) = \sigma^2$, $E(s) = E(\sqrt{s^2}) \le \sqrt{Es^2} = \sigma$, by Jensen

Strong Law of Large Numbers: $\overline{X_n} \to \mu$ with probability 1, as $n \to \infty$

Weak Law of Large Numbers:

for any $\epsilon > 0$,

$P(|\overline{X_n} - \mu| \ge \epsilon) \to 0$ as $n \to \infty$, proof using Chebyshev

Central Limit Theorem

$T_n = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \to N(0, 1)$, for large n

$\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n})$