Table of Contents

Chapter 1: Probability

- Intersection - probability that both A and B occur
- Complement - A^c event that A does not occur, all events in the sample space that are not A
- Disjoint - A and C are disjoint if $A \cap C = \emptyset$
- Probability Axioms: 1) $P(\Omega) = 1$, 2) If $A \subset \Omega$ then $P(A) \geq 0$ 3) If A, B disjoint then $P(A \cup B) = P(A) + P(B)$
- Addition Law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Permutation: ordered arrangement of objects
- Binomial coefficients: $(a+b)^n = \sum_{k=0}^{n} a^k b^{n-k}$
- # of ways n objects can be grouped into r classes with n_I in the i^{th} class: $\binom{n}{n_1 n_2 \ldots n_r} = \frac{n!}{n_1! n_2! \ldots n_r!}$
- Bayes, multiplication law: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, $P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum P(A|B_i)P(B_i)}$
- Law of total probability: $P(A) = \sum P(A|B_i)P(B_i)$
- Independence for sets: $P(A \cap B) = P(A)P(B)$. Mutual independence implies pairwise independence

Chapter 2: Random Variables

- CDF: $F(x) = P(X \leq x)$

Discrete Mass Functions

- Bernoulli: $p\left(x\right)=\begin{Bmatrix} p^{x}\left(1-p\right)^{1-x}, & if\, x=0\, or\, x=1 \\ 0, & otherwise \end{Bmatrix}, \, 0\leq p\leq 1$
- All moments = p
- Binomial: $p(k)=\binom{n}{k}p^{k}(1-p)^{n-k}$
  $n\in\mathbb{N}, k=0,1,\ldots,n, \, 0\leq p\leq 1$
- $Bin(n,p)=\sum_{i=1}^{n}X_i$, for X Bern(P)
  - $E(X)=np, Var(x)=np(1-p)$
- Geometric: $p(k)=p(1-p)^{k-1}, \quad k=\mathbb{N}$
  - $$E(X)=\frac{1}{p}$$ $$\mathrm{Var}(X)=\frac{1-p}{p^2}$$
- Negative Binomial: $P(X=k)=\binom{k-1}{r-1}p^{r}(1-p)^{k-r} \qquad ,0\leq p\leq 1, k=r,r+1,\ldots,r=1,2,\ldots,k$
- Hypergeometric
  - *n*: population size; *n*∈N
  - *r*: successes in population; *r*∈{0,1,...,*n*}
  - *m*: number drawn from population; *m*∈{0,1,...,*n*}
  - *X*: number of successes in drawn group
  - $P(X=k)=\frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}\ \max(0,m+r-n)\leq k\leq\min(r,m)\, 0\leq p(k)\leq 1$
- Poisson: $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}\ k=0,1,2,3,\ldots,\ \lambda>0$
  - $E(X)=\lambda, Var(X)=\lambda$

Continuous Density Functions

- Uniform: $f(x)=\begin{cases} 1/(b-a) & a\leq x\leq b \\ 0 & x<a\, or\, x>b \end{cases}, x\in[a,b]\, , a<b\, , f:\mathbb{R}\mapsto[0,\infty)$
  - $E(X)=\frac{1}{2}(a+b)$
  - $Var(X)=\frac{1}{12}(b-a)^2$
- Exponential $f(x)=\begin{cases} \lambda e^{-\lambda x}, & x\geq 0 \\ 0, & x<0 \end{cases}, \lambda>0$
  - $E(X)=\frac{1}{\lambda}$
  - $Var(X)=\frac{1}{\lambda^2}$
- Normal: $f(x\mid\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, f:\mathbb{R}\to(0,\infty), \mu\in\mathbb{R}, \sigma>0$
- Gamma:
  - $\Gamma(x)=\int_0^\infty u^{x-1}e^{-u}\,du, x>0$
  - $g(t\mid\alpha,\lambda)=\begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-\lambda t}, t\geq 0 \\ 0, t<0 \end{cases}$
  - $g:\mathbb{R}\to[0,\infty), \alpha>0, \lambda>0$
  - $$E(X)=\frac{\alpha}{\lambda}$$ $$\mathrm{Var}(X)=\frac{\alpha}{\lambda^2}$$

- Beta: $f(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1}, 0 \le u \le 1,\ a, b > 0$
  - $$E(X) = \frac{a}{a+b}$$
    $$\mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$
- Cauchy: $f_Z(z) = \frac{1}{\pi(z^2+1)}$ for $z \in (-\infty, \infty)$

Functions of RVs

- If $X \sim N(\mu, \sigma^2) \mathrm{and} Y = aX + b,\ \text{then } Y \sim N(a\mu + b, a^2\sigma^2).$
- **Change of variables for function of RV** $f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy} g^{-1}(y)\right|$
  - Note if you do with CDFs, if z = x/2 then P(Z < z) = P(X/2 < z) = P(X < 2z) = $F_Z(2z)$
  - Then $f_z(2z) * 2$, which is the same thing as taking the Jacobian from the beginning
- Universality of uniform: Z = F(X) for $Z \sim U(0,1)$. $X = F^{-1}(U)$ then $F(x) = P(U \le F(x))$
- If we sum n independent Gam(α) random variables we get a Gam(nα) random variable.

Chapter 3: Joint Distributions

- Marginal frequency: $p_X(x) = \sum_i p(x, y_i)$
- Joint density, $f(x,y)$, then $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v) dv\, du$
- Marginal cdf $F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u,y) dy\, du$, marginal density $f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy$
- Copula: a joint CDF of RVs that have uniform marginal distributions
- Random variables $X_1, X_2, \ldots, X_n$ are independent if the joint CDF factors into the product of marginals: $F(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \ldots F_{X_n}(x_n)$
- Conditional density: $f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$
- Convolution: X, Y RVs, Z = X + Y, then density of Z found by noting $Z = z$ when $X = x$ and $Y = z - x$. Therefore $f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x) dx$ is the convolution of f_X, f_Y

Variable Transformation

- $u = g_1(x,y),\ v = g_2(x,y)$ for X, Y jointly distributed RVs
- Invert the transformation to $x = h_1(u,v),\ y = h_2(u,v)$
- Take jacobian wrt u and v for x and y ie. $\frac{\partial x}{\partial u}$
- Then change of variables = $f_{UV}(u,v) = f_{XY}(h_1(u,v), h_2(u,v))|J|$

Chapter 4: Expected Values

- Discrete: $E(X) = \sum_i x_i p(x_i)$
- Continuous: $E(X) = \int_{-\infty}^{\infty} x f(x) dx$ provided the expectation of |x| exists.
  - When we get the marginal from a joint, even if x is bound in terms of y, the bounds of the marginal are independent of y. This is why the expectation is integrated over the whole domain.
- E(X) is a constant. E(X|Y) is a function of Y, and thus a RV.
- Functions of RVs: $E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$ - do not need the pdf of g(x) for expectation
- For independent X and Y, E(XY) = E(X)E(Y).
- Linearity of expectation - constants and coefficients can be pulled out of expectation

- Variance
    - $\operatorname{Var}(X)=\sum_i\left(x_i-\mu\right)^2 p\left(x_i\right)$
    - $\operatorname{Var}(X)=E\left\{[X-E(X)]^2\right\}$
    - $\operatorname{Var}(X)=E\left(X^2\right)-[E(X)]^2$
    - $\operatorname{Var}(X)=Cov(X,X)$
    - $\operatorname{Cov}(X, Y)=E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]=E(X Y)-E(X) E(Y)$. If X,Y independent, Cov = 0
    - $\operatorname{Cov}(aX, bY)=abCov(X,Y)$
    - $\begin{aligned} \operatorname{Var}(X+Y) &= \operatorname{Cov}(X+Y, X+Y) \\ &= \operatorname{Var}(X)+\operatorname{Var}(Y)+2 \operatorname{Cov}(X, Y)\end{aligned}$
- We can standardize a RV X: $\frac{X-E(X)}{\sqrt{Var(X)}}$. Has mean 0, SD 1.
- $Corr(X,Y)=Cov(\frac{X-E(X)}{\sqrt{Var(X)}},\frac{Y-E(Y)}{\sqrt{Var(Y)}})=\frac{\sigma_X Y}{\sigma_X\sigma_Y}$
- Conditional Expectation: $E(Y|X=x)=\int y f_{Y|X}(y|x)dy$
- Law of total expectation: E(Y) = E(E(Y|X)), $\operatorname{Var}(Y)=\operatorname{Var}[E(Y|X)]+E[\operatorname{Var}(Y|X)]$

Markov Inequality

- X RV with domain geq 0, then $P(X\geq t)\leq\frac{E(X)}{t}$
- This result says that the probability that X is much bigger than E(X) is small.

Chebyshev's Inequality

- $P(|X-\mu|>t)\leq\frac{\sigma^2}{t^2}$
- Plug in t = K x sigma to get alternate form

Moment Generating Functions

- Uniquely determines a probability distribution - same MGF means same distribution
- $M(t)=E(e^{tX})=\int_{-\infty}^{\infty} e^{tx} f(x)\,dx$
- The rth moment: $M^{(r)}(0)=E(X^r)$
- $\text{If } X \text{ has the mgf } M_X(t) \text{ and } Y=a+bX, \text{ then } Y \text{ has the mgf } M_Y(t)=e^{at}M_X(bt)$
- If X, Y independent and Z= X + Y, Mz(t) = Mx(t)My(t)

delta Method

- For $Y=g(x), E(X)=\mu, Var(X)=\sigma^2$
- $Y\approx g(\mu)+g'(\mu)(X-\mu)$
- $\sigma_Y^2\approx\sigma_X^2\left[g'\left(\mu_X\right)\right]^2$
- $E(Y)\approx g\left(\mu_X\right)+\frac{1}{2}\sigma_X^2 g''\left(\mu_X\right)$

Chapter 5: Limit Theorems

Weak Law of Large Numbers (Convergence in probability)

- $X_1...X_i\sim iid$, $E(X_i)=\mu$ and $Var(X_i)=\sigma^2$. Let $X_n=n^{-1}\sum_{i=1}^n X_i$ and
- $lim_{n\to\infty}P(|\bar{X}_n-\mu|>\epsilon)=0$
    - Averages settle down to their corresponding expectations
- Need to have constant variance to apply WLLN -> otherwise must normalize across X_is

- Convergence in probability - the probability goes to zero, but we cannot say what happens to each point in the distribution. We are interested in the output of the CDF in limit.
- Weak is for a given epsilon, holding that constant, we will converge in probability to zero. For strong, we say for all epsilon it will eventually converge.

Convergence in Distribution

- X_1, X_2 sequence of RVs with CDFs $F_1, F_2 \dots$. Let X be RV with distribution function F. Say X_n converges in distribution to X if:
- $\lim_{n\to\infty} F_n(x) = F(x)$ at every point for continuous F
- MGFs usually used to show this property.

Almost Sure Convergence

- The difference between a point and alpha is greater than alpha a finite number of times. In the infinite limit, there is some point at which the difference is always less than epsilon.
- $\varepsilon > 0, |Z_n - \alpha| > \varepsilon$

Continuity Theorem

- F_n be a sequence of CDFs with MGF M_n. F CDF with MGF M.
- If $M_n(t) \to M(t)$ for all t then $F_n(x) \to F(X)$ for all points of F

Central Limit Thm

- For X_i iid, $E(X) = \mu$, $\ Var(X) = \sigma^2$
- $\lim_{n\to\infty} \Pr\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le x\right) = \Phi(x)$
  - Averages have a nearly normal distribution
- Book defn: X_1, X_2 sequence of RVs with mean 0 and variance σ^2 common CDF F and MGF M. $S_n = \Sigma_{i=1}^n X_i$, $\lim_{n\to\infty} P(\frac{S_n}{\sigma\sqrt(n)} \le x) = \Phi(x)$ on $-\infty < x < \infty$

Chapter 6: Distributions Derived from the Normal

Chi Square Distribution

- U = Z^2 for $Z \sim N(0,1)$, then $U \sim \chi_1^2$
- Note that the distribution is equivalent to $\chi_1^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$, $\chi_n^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2})$
- If U1, U2, ... ,Un are independent chi-square random variables with 1 degree of freedom, the distribution of V = U1 + U2 + ⋯ + Un is called the chi-square distribution with n degrees of freedom and is denoted by χ_n^2

t-distribution

- $Z \sim N(0,1)$ and $U \sim \chi_n^2$ for Z, U independent, then
- $\frac{Z}{\sqrt{\frac{U}{n}}}$ is a t-distribution with n degrees of freedom
- Density function of $f(t) = \frac{\Gamma[\frac{1}{2}(n+1)]}{\sqrt{n\pi}\Gamma(n/2)}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$ on $t \in \mathbb{R}$
- Notice $t_1 \sim Cauchy$

F Distribution

- U, V independent chi-square RVs wit m and n respective DoF
- $W = \frac{U/M}{V/n}$ is F with m and n DoF, ie. $F_{m,n}$
- $f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} w^{m/2-1} \left(1 + \frac{m}{n} w\right)^{-(m+n)/2}$ for $w \geq 0$

Sample Statistics

- For $X_1, \ldots, X_n$ iid **normals**, we sometimes refer to them as a sample from a normal distribution.
- $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ = sample mean
- $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ = sample variance
- Note that $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$
- The RV X-bar and the vector of RVs $(X_1 - \bar{X}, \ldots)$ are independent
- $\bar{X}$ and s^2 are independently distributed
- The distribution of $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$
- Important: $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$. We have one unknown value μ, but from a known distribution allowing us to make a CI with the t distribution
- Note: normalizing - always subtracting off the mean and dividing by the standard deviation of that variable.
- Var($\bar{X}_n$) = $\frac{\sigma^2}{n}$, then $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$