

Contents

1	Distributions	3
1.1	Bernoulli	3
1.2	Poisson	3
1.3	Geometric	3
1.4	Binomial	3
1.5	Negative Binomial	4
1.6	Hypergeometric	4
1.7	Exponential	4
1.8	Normal	4
1.9	Uniform	5
1.9.1	Uniform Continuous	5
1.9.2	Uniform Discrete	5
1.10	Gamma Distribution	5
1.11	Beta Distribution	5
1.12	Mgfs	6
1.13	Location and Scale Families	6
2	Transformations	8
2.1	Integration by Substitution of a New Variable	8
2.2	Application to Probability Density Functions	8
2.3	Constraints	9
3	Notes 13	10
3.1	Multinomial Distribution	10
3.2	Exponential Families	10
3.3	Conditional Probability	10
4	Notes 14	11
4.1	Convolution	11
4.2	Sum of Two Independent Poissons	11
4.3	Jacobian	11
4.4	Functions of Independent Random Variables	12
4.5	Ratio of Two Independent Normals	12
4.6	Sum of Two Independent Random Variables	12

5	Notes 15	13
5.1	Conditional Expectation and Variance	13
5.1.1	Expectation	13
5.1.2	Variance	13
5.2	Hierarchical Model Example	14
5.3	Covariance and Correlation	14
5.4	Linear Combinations	15
5.5	Standard Bivariate Normal	15
5.6	Bivariate Normal	15
5.7	Multivariate Distributions	15
5.8	Marginals and Conditionals	16
5.9	Multivariate Independence	16
5.10	Multinomial	16
6	Inequalities	17
6.1	Chebychev Inequality	17
6.2	L^p Spaces	17
6.3	Triangle Inequality	17
6.4	Jensen's Inequality	17
6.4.1	Convex Functions	18
6.5	Young's Inequality	18
6.6	Holder's Inequality	18
6.6.1	Cauchy-Schwartz Inequality	18
6.6.2	Lyapunov's Inequality	18
6.7	Minkowski's Inequality	18
7	Order Statistics	19
7.1	Min and Max	19
7.1.1	Distribution of the Maximum	19
7.1.2	Distribution of the Minimum	19
7.2	Order Statistics	19
7.3	r^{th} Order Statistic	20
7.3.1	Example with U(0,1)	20
7.4	Distribution of the Median	20
7.5	Joint Distribution of $X_{(i)}$ and $X_{(j)}$	21
7.5.1	Example with U(0,1)	21
7.6	Joint Distribution of all Order Statistics	21
7.7	Distribution of the Range	21
7.7.1	Range Example	22

Chapter 1

Distributions

1.1 Bernoulli

$$\begin{aligned} p^x(1-p)^{1-x} \quad x = 0, 1 \\ EX = p \quad VarX = p(1-p) \end{aligned}$$

1.2 Poisson

Expresses the probability of a given number of events occurring during a fixed interval of time or space if these events occur with a known constant rate independently of the time since the last event.

$$\begin{aligned} \frac{\lambda^x e^{-\lambda}}{x!} \\ EX = \lambda \quad VarX = \lambda \end{aligned}$$

1.3 Geometric

The probability distribution of the number X of bernoulli trials needed to get one success.

$$\begin{aligned} (1-p)^{x-1}p \quad x = 1, \dots \\ 1 - (1-p)^x \\ EX = \frac{1}{p} \quad VarX = \frac{1-p}{p^2} \end{aligned}$$

1.4 Binomial

distribution of the number of successes in a sequence of n independent bernoulli trials.

$$\begin{aligned} \binom{n}{x} p^x (1-p)^{n-x} \\ EX = np \quad VarX = np(1-p) \end{aligned}$$

1.5 Negative Binomial

number of successes in a sequence of iid bernoulli trials before a specified number of failures (r).

$$\binom{k+r-1}{k} (1-p)^r p^k$$

$$\text{Negative Binomial Mgf} \left(\frac{p}{1 - (1-p)e^t} \right)^r$$

$$EX = \frac{r(1-p)}{p} \quad VarX = \frac{r(1-p)}{p^2}$$

1.6 Hypergeometric

The result of each draw (the elements of the population being sampled) can be classified into two mutually exclusive categories ie pass/fail.

The probability of a success changes on each draw, as each draw decreases the population (sampling without replacement from a finite population)

N=population size

M=number of success in the population

K= number of draws

x=number of observed successes

$$\frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}}$$

$$EX = \frac{KM}{N} \quad VarX = \frac{KM}{N} \frac{(N-M)(N-K)}{N(N-1)}$$

1.7 Exponential

describes the waiting time between Poisson events, Memoryless

$$\lambda e^{-\lambda x} \quad 0 \leq x < \infty \quad \lambda > 0$$

$$1 - e^{-\lambda x}$$

$$EX = \frac{1}{\lambda} \quad VarX = \frac{1}{\lambda^2}$$

1.8 Normal

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

$$EX = \mu \quad VarX = \sigma^2$$

1.9 Uniform

1.9.1 Uniform Continuous

symmetric probability distribution where all intervals of the same length are equally probable.

$$\frac{1}{b-a}$$

$$EX = \frac{b+a}{2} \quad VarX = \frac{(b-a)^2}{12}$$

1.9.2 Uniform Discrete

a symmetric probability distribution where a finite number of values are equally likely to be observed. Every one of n values has an equal probability $1/n$.

$$\frac{1}{n} \quad x = 1, \dots, N \quad N = 1, \dots$$

$$EX = \frac{N+1}{2} \quad VarX = \frac{(N+1)(N-1)}{12}$$

1.10 Gamma Distribution

Gamma Function: $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha) \quad \alpha > 0$$

$$\Gamma(n) = (n-1)! \quad n \in \mathbb{Z}$$

$$\Gamma(1/2) = \sqrt{\pi}$$

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

α is the shape parameter, influences the peakedness of the distribution

β is the scale parameter, influences the spread of the distribution

$$EX^v = \frac{\beta^v \Gamma(v+\alpha)}{\Gamma(\alpha)}$$

$$\Gamma(\alpha+v) = \int_0^\infty x^{v+\alpha-1} e^{-x} dx$$

$$EX = \alpha\beta \text{ alternatively } a/\lambda \quad \lambda = 1/\beta$$

$$Var(X) = \alpha\beta^2 \text{ alternatively } a/\lambda^2$$

$$\int_0^\infty e^{-x^2/2} dz = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}$$

$$\int_0^\infty x^2 e^{-x^2} \text{ is the same}$$

1.11 Beta Distribution

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad 0 \leq x \leq 1$$

Beta Function: $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

$$EX^n = \frac{B(\alpha + n, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha + n)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + n)\Gamma(\alpha)}$$

$$E(X) = \frac{a}{a + b}$$

$$Var(X) = \frac{ab}{(a + b)^2(a + b + 1)}$$

$$beta(1, 1) = U(0, 1)$$

1.12 Mgfs

Specific Characteristic Functions

	<u>mgf</u>	<u>cf</u>
Bernoulli(p)	$pe^t + q$	$pe^{it} + q$
Binomial(n,p)	$(pe^t + q)^n$	$(pe^{it} + q)^n$
Poisson(λ)	$e^{\lambda(e^t - 1)}$	$e^{\lambda(e^{it} - 1)}$
Geometric(p)	$pe^t / (1 - qe^t)$	$pe^{it} / (1 - qe^{it})$
Negbin(n, p)	$\left[\frac{pe^t}{1 - qe^t} \right]^n$	$\left[\frac{pe^{it}}{1 - qe^{it}} \right]^n$
Uniform(a, b)	$\frac{e^{tb} - e^{ta}}{t(b - a)}$	$\frac{e^{itb} - e^{ita}}{it(b - a)}$
Normal(μ, σ^2)	$e^{t\mu + \frac{1}{2}\sigma^2 t^2}$	$e^{it\mu - \frac{1}{2}\sigma^2 t^2}$
Exponential(λ)	$\frac{\lambda}{\lambda - t} = (1 - \frac{t}{\lambda})^{-1}$	$(1 - \frac{it}{\lambda})^{-1}$
Gamma(a, λ)	$(1 - \frac{t}{\lambda})^{-a}$	$(1 - \frac{it}{\lambda})^{-a}$

1.13 Location and Scale Families

$$f_{\mu, \sigma}(x) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

$\mu \in \mathbb{R}$ $\sigma > 0$ is a location scale family

If $\mu = 0$ scale family

If $\sigma = 1$ location family

Properties: Let $Z \sim f(z)$ and $X = \sigma Z + \mu$ Then

X has pdf $f_{\mu,\sigma}$

$$E(X) = \sigma E(Z) + \mu \quad \text{Var}(X) = \sigma^2 \text{Var}(Z)$$

Chapter 2

Transformations

2.1 Integration by Substitution of a New Variable

General Case

$$\int_a^b f(x) \, dx$$

Choose a transformation function $y(x)$

Note its inverse $x(y)$

Replace the limits by $y(a)$ and $y(b)$

Replace $f(x)$ by $f(x(y))$

Replace dx by $\frac{dx}{dy} dy$ (derivative of inverse)

Result is $\int_{y(a)}^{y(b)} f(x(y)) \frac{dx}{dy} dy$

Example

$$\int_0^{\sqrt{\frac{\pi}{2}}} 2x \cos(x^2) \, dx$$

$$y(x) = x^2$$

$$x(y) = \sqrt{y}$$

$$0 \text{ and } \frac{\pi}{2}$$

$$2\sqrt{y} \cos(y)$$

$$\frac{1}{2\sqrt{y}} dy$$

$$\int_0^{\frac{\pi}{2}} \cos(y) \, dy = 1$$

2.2 Application to Probability Density Functions

$$\int_a^b f(x) \, dx = \int_{y(a)}^{y(b)} f(x(y)) \frac{dx}{dy} dy$$

$$P(a \leq X < b) = \int_a^b f(x) \, dx$$

Any function of an r.v. is an r.v. itself and if y is taken as some transformation function, $y(X)$ will be a derived r.v.

Let $Y = y(X)$

Notice that if $X = a$ the derived r.v. $Y = y(a)$ and if $X = b$, $Y = y(b)$

If $a \leq X < b$ then $y(a) \leq Y < y(b)$ and:

$P(y(a) \leq Y < y(b)) = P(a \leq X < b)$ Therefore:

$$P(y(a) \leq Y < y(b)) = P(a \leq X < b) = \int_a^b f(x) \, dx = \int_{y(a)}^{y(b)} f(x(y)) \frac{dx}{dy} dy$$

Let $g(y) = f(x(y)) \frac{dx}{dy}$ Then:

$$P(y(a) \leq Y < y(b)) = \int_{y(a)}^{y(b)} g(y) \, dy$$

Thus $g(y)$ is the pdf associated with Y

2.3 Constraints

The function $f(x)$ is transformed by $y(x)$ into $g(y)$

Throughout the useful range of x, both $y(x)$ and it's inverse $x(y)$ must be defined and single valued

Throughout this range, $\frac{dx}{dy}$ must be defined and must not change signs.

The transformation function $y(x)$ must be monotonic

$$g(y) = f(x(y)) \left| \frac{dx}{dy} \right|$$

Chapter 3

Notes 13

3.1 Multinomial Distribution

$$p(s_1, s_2, \dots, s_k) = \frac{n!}{s_1! s_2! \dots s_k!} p_1^{s_1} p_2^{s_2} \dots p_k^{s_k}$$

where $\sum_{i=1}^k s_i = n$ and $\sum_{i=1}^k p_i = 1$

3.2 Exponential Families

A family of pdfs or pmfs is called an exponential family if it can be expressed as

$$f(x|\theta) = h(x)c(\theta) \exp \left(\sum_{i=1}^k w_i(\theta) t_i(x) \right)$$

3.3 Conditional Probability

Discrete:

$$\begin{aligned} P\{Y = y|X = x\} &= \frac{P(X = x, Y = y)}{P(X = x)} \\ &= \frac{f_{X,Y}(x, y)}{f_X(x)} \\ &= f_{Y|X}(y|x) \end{aligned}$$

Can rewrite as:

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

Continuous: $f(y|x) = \frac{f(x, y)}{f(x)}$

Chapter 4

Notes 14

4.1 Convolution

If X and Y are independent continuous r.v.s with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $Z = X + Y$ is:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z - w) dw$$

4.2 Sum of Two Independent Poissons

$$X \sim \text{Pois}(\lambda_1), Y \sim \text{Pois}(\lambda_2)$$

$$U = X + Y \quad V = Y$$

$$X = U - V \quad Y = V$$

Joint PMF of U and V is:

$$f_{U,V}(u, v) = f_{X,Y}(u - v, v) = \frac{e^{-\lambda_1} \lambda_1^{u-v}}{(u-v)!} \frac{e^{-\lambda_2} \lambda_2^v}{v!}$$

The distribution of $U = X + Y$ is the marginal:

$$\begin{aligned} f_U(u) &= \sum_{v=0}^u \frac{e^{-\lambda_1} \lambda_1^{u-v}}{(u-v)!} \frac{e^{-\lambda_2} \lambda_2^v}{v!} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)}}{u!} \sum_{v=0}^u \binom{u}{v} \lambda_1^{u-v} \lambda_2^v \end{aligned}$$

Because of the binomial theorem

$$= \frac{e^{-(\lambda_1 + \lambda_2)}}{u!} (\lambda_1 + \lambda_2)^u$$

$$U \sim \text{Pois}(\lambda_1 + \lambda_2)$$

4.3 Jacobian

$J(u, v)$ is the Jacobian of the transformation $(x, y) \rightarrow (u, v)$ given by:

$$J(u, v) = \frac{\partial(x, y)}{\partial(u, v)} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$

4.4 Functions of Independent Random Variables

Let X and Y be independent r.v.s

Let $g : \mathbb{R} \rightarrow \mathbb{R}$ and $h : \mathbb{R} \rightarrow \mathbb{R}$ be functions

Then the r.v.s $U = g(X)$ and $V = h(Y)$ are independent

4.5 Ratio of Two Independent Normals

Let $X \sim N(0, 1)$ and $Y \sim N(0, 1)$

The ratio X/Y has the Cauchy distribution

Let $U = X/Y$ and $V = Y$ Then $X = UV$ and $Y = V$ $J(u, v) = v$

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

$$f_{U,V}(uv, v) = \frac{1}{2\pi} e^{-[(uv)^2+v^2]/2} * |v| = \frac{|v|}{2\pi} e^{-(u^2+1)v^2/2}$$

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} f_{UV}(u, v) dv = 2 \int_0^{\infty} \frac{v}{2\pi} e^{-(u^2+1)v^2/2} dv \\ &= \frac{1}{\pi} \int_0^{\infty} e^{-(u^2+1)z} dz = \frac{1}{\pi(u^2+1)} \end{aligned}$$

4.6 Sum of Two Independent Random Variables

Suppose X and Y are independent, find distribution of $Z = X + Y$

In general: $F_Z(z) = P(X + Y \leq z) = P(\{(x, y) \text{ such that } x + y \leq z\})$

Approaches:

- Bivariate transformation method (continuous and discrete)
- Discrete convolution:
 $f_Z(z) = \sum_{x+y=z} f_X(x) f_Y(y) = \sum_x f_X(x) f_Y(z-x)$
- Continuous convolution
- Mgf/cf method (continuous and discrete)
 $\phi_Z\theta = \phi_X(\theta) \phi_Y(\theta)$
 $Z = X - Y \quad \phi_Z\theta = \phi_X(\theta) \phi_Y(-\theta)$

Chapter 5

Notes 15

5.1 Conditional Expectation and Variance

5.1.1 Expectation

For two r.v.s X and Y with conditional pdf $f_{Y|X}(y|x)$ the conditional expectation of $g(Y)$ given $X = x$ is:

If discrete:

$$E[g(Y)|X = x] = \sum_y g(y)f_{Y|X}(y|x)$$

If continuous:

$$h(x) = E[g(Y)|x] = \int_{-\infty}^{\infty} g(y)f_{Y|X}(y|x) dy$$

$$h(X) = E[g(Y)|X]$$

Iterative Expectation Formula

$$EX = E(E(X|Y))$$

5.1.2 Variance

$$Var[g(Y)] = E[g(y) - E(g(Y))]^2$$

$$VarX = E(Var(X|Y)) + Var(E(X|Y))$$

$$Var(g(Y)|X) = E\{[g(Y) - E(g(Y)|X)]^2|X\}$$

where both expectations are taken with respect to $f_{Y|X}(y)$

- $E(Var(X|Y)) = E\{[X - E(X|Y)]^2\}$
- $Var(E(X|Y)) = E\{[E(X|Y) - EX]^2\}$

$$Var(aX) = a^2Var(x)$$

5.2 Hierarchical Model Example

An insect lays eggs according to a $Pois(\lambda)$ and each egg survives with probability p , independent

Find average eggs surviving

$Y \sim Pois(\lambda)$ $X|Y \sim Binom(Y, p)$ Where X is total eggs surviving

$$P(X = x) = \sum_{y=0}^{\infty} P(X = x, Y = y)$$

$$= \sum_{y=0}^{\infty} P(X = x|Y = y)P(Y = y) \text{ using conditional expectation}$$

$$= \sum_{y=x}^{\infty} \left[\binom{y}{x} p^x (1-p)^{y-x} \right] \left[\frac{e^{-\lambda} \lambda^y}{y!} \right]$$

$$= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{((1-p)\lambda)^{y-x}}{(y-x)!}$$

$$\text{Since } e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

$$\text{Let } x = (1-p)\lambda \quad k = y - x$$

$$\frac{(\lambda p)^x e^{-\lambda}}{x!} e^{(1-p)\lambda}$$

$$= \frac{(\lambda p)^x}{x!} e^{-\lambda p}$$

$$\text{So } X \sim Pois(\lambda p) \quad EX = \lambda p$$

Using iterative expectation formula:

$$EX = E(E(X|Y)) = E(Yp) = \lambda p$$

5.3 Covariance and Correlation

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}$$

$$\text{Correlation} = \rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var X} \sqrt{Var Y}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$= E \left[\left(\frac{X - \mu_X}{\sigma_X} \right) \left(\frac{Y - \mu_Y}{\sigma_Y} \right) \right]$$

X and Y are uncorrelated iff:

$$Cov(X, Y) = 0 \text{ or equivalently } \rho_{XY} = 0$$

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

If X and Y are independent and $Cov(X, Y)$ exists, then $Cov(X, Y) = 0$

If X and Y are uncorrelated this does not imply independence.

5.4 Linear Combinations

$$\text{Cov}(aX + bY, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z)$$

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

$$\text{Corr}(aX + b, cY + d) = \frac{ac}{|ac|}\text{Corr}(X, Y)$$

5.5 Standard Bivariate Normal

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right]$$

Both X and Y have marginal distributions are $N(0, 1)$

Correlation of X and Y is ρ

Conditional Distribution are normal:

$$Y|X \sim N(\rho X, 1 - \rho^2) \quad X|Y \sim N(\rho Y, 1 - \rho^2)$$

The means are the regression lines of Y on X and X on Y respectively.

5.6 Bivariate Normal

Let \tilde{X} and \tilde{Y} have a standard bivariate normal distribution with correlation ρ

$$\text{Let } X = \mu_X + \sigma_X \tilde{X} \quad Y = \mu_Y + \sigma_Y \tilde{Y}$$

Then (X, Y) has the bivariate normal density:

$$f_{XY}(x, y) = \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}\right)^{-1} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

Marginal distributions: $N(\mu_X, \sigma_X^2) \quad N(\mu_Y, \sigma_Y^2)$

$$\text{Corr}(X, Y) = \rho$$

Conditional distributions are normal:

$$Y|X \sim N[\mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X), \sigma_Y^2(1 - \rho^2)]$$

Distribution of $aX + bY$ is:

$$N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$$

5.7 Multivariate Distributions

$$\mathbf{X} = (X_1, X_2, \dots, X_n)$$

If \mathbf{X} is discrete then:

$$P(\mathbf{X} \in A) = \sum_{\mathbf{X} \in A} f(\mathbf{X})$$

where $f(\mathbf{X})$ is the joint pmf

If \mathbf{X} is continuous then:

$$P(\mathbf{X} \in A) = \int \cdots \int_A f(x_1, \dots, x_n) dx_1, \dots, dx_n$$

5.8 Marginals and Conditionals

The **marginal** pdf or pmf of any subset of coordinates is found by integrating or summing the joint pdf or pmf over all possible values of the other coordinates.

The **conditional** pdf or pmf of a subset of coordinates given the values of the remaining coordinates is found by dividing the full joint pdf or pmf by the joint pdf or pmf of the remaining variables.

5.9 Multivariate Independence

Independent Random Vectors:

Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be random vectors with joint pdf or pmf $f(\mathbf{X}_1, \dots, \mathbf{X}_n)$

Let $f\mathbf{X}_j(x_j)$ be the marginal pdf or pmf of \mathbf{X}_j .

Then $\mathbf{X}_1, \dots, \mathbf{X}_n$ are **mutually independent** random vectors if:

$$\forall (\mathbf{X}_1, \dots, \mathbf{X}_n): f(\mathbf{X}_1, \dots, \mathbf{X}_n) = \prod_{j=1}^n f\mathbf{X}_j(x_j)$$

5.10 Multinomial

Let n and m be positive integers and let p_1, \dots, p_n be probabilities summing to one. Then the random vector (X_1, \dots, X_n) has a multinomial distribution with m trials and cell probabilities p_1, \dots, p_n if its joint pmf is:

$$\begin{aligned} f(x_1, \dots, x_n) &= \binom{m}{x_1, \dots, x_n} p_1^{x_1} \cdots p_n^{x_n} \\ &= \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n} \\ &= m! \prod_{j=1}^n \frac{p_j^{x_j}}{x_j!} \end{aligned}$$

for $x_1 = 0, \dots, m$ $i = 1, \dots, n$ $x_1 + \cdots + x_n = m$

Chapter 6

Inequalities

6.1 Chebychev Inequality

$$P[g(X) \geq r] \leq \frac{E[g(X)]}{r}$$

If X is nonnegative and g is a positive non-decreasing function then:

$$P\{X \geq a\} \leq \frac{E[g(X)]}{g(a)}$$

Special Cases:

$$X \geq 0 \quad P\{X \geq a\} \leq \frac{E(e^{tX})}{e^{ta}}$$

6.2 L^p Spaces

L^p Space consists of all r.v.s whose p^{th} absolute power is integrable,
 $E(|X|^p) < \infty$

6.3 Triangle Inequality

$$|a + b| \leq |a| + |b|$$

6.4 Jensen's Inequality

Let $X \in L^1$ and $g(x)$ be a convex function where $E[g(X)]$ exists. Then:

$$E[g(X)] \geq g[EX]$$

with equality iff for every line $a + bx$ tangent to $g(x)$ at

$$x = EX, P[g(X) = a + bX] = 1$$

direction of inequality is reversed if g is concave

6.4.1 Convex Functions

A function $g : I \rightarrow \mathbb{R}$ is convex for any $\lambda \in [0, 1]$ and any points x and y in I
 $g[\lambda x + (1 - \lambda)y] \leq \lambda g(x) + (1 - \lambda)g(y)$
 A differentiable function g is convex iff it lays above all tangents.
 A twice differentiable function g is convex iff its second derivative is non-negative
 concave if $-g$ is convex on I

6.5 Young's Inequality

Let $a, b > 0$ and $p, q > 1$ with $1/p + 1/q = 1$ Then:

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab$$

with equality iff $a^p = b^q$

6.6 Holder's Inequality

Suppose $X \in L^p, Y \in L^q$ where $p, q > 1$ and $1/p + 1/q = 1$ Then:
 $E[|XY|] \leq [E|X|^p]^{1/p} E[|Y|^q]^{1/q}$
 with equality if $X^p = cY^q$ for some $c \in \mathbb{R}$

6.6.1 Cauchy-Schwartz Inequality

corollary of Holders where $p = q = 2$
 $E[|XY|] \leq [E|X|^2]^{1/2} E[|Y|^2]^{1/2} = \sqrt{E[X^2]E[Y^2]}$
 with equality if $X = cY$

Application of Cauchy-Schwartz

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

Then $|\rho| \leq 1$ with equality iff $Y - \mu_Y = c(X - \mu_X)$

6.6.2 Lyapunov's Inequality

corollary of Holders
 for $1 \leq r \leq s$ and $X \in L^s$
 $E[|X|^r]^{1/r} \leq E[|X|^s]^{1/s}$

6.7 Minkowski's Inequality

Suppose $X, Y \in L^p, p \geq 1$ Then $(X + Y) \in L^p$ and
 $E|X + Y|^p]^{1/p} \leq [E|X|^p]^{1/p} + [E|Y|^p]^{1/p}$

Chapter 7

Order Statistics

7.1 Min and Max

7.1.1 Distribution of the Maximum

The cdf of $Z = \max(Y_1, \dots, Y_n)$ is

$$\begin{aligned} F_Z(z) &= P\{Z \leq z\} \\ &= P\{Y_1 \leq z, Y_2 \leq z, \dots, Y_n \leq z\} \\ &= \prod_{j=1}^n P\{Y_j \leq z\} \text{ indep} \\ &= F_Y(z)^n \text{ ident. distrib} \end{aligned}$$

Thus the pmf is:

$$f_Z(z) = nF_Y(z)^{n-1}f_Y(z)$$

7.1.2 Distribution of the Minimum

$$W = \min(Y_1, \dots, Y_n)$$

$$F_W(w) = 1 - (1 - F_Y(w))^n$$

$$f_W(w) = n(1 - F_Y(w))^{n-1}f_Y(w)$$

7.2 Order Statistics

Let Y_1, Y_2, \dots, Y_n be iid with pdf $f_Y(x)$

Order the observations:

$$Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$$

The $Y_{(i)}$ are called order statistics. Minimum is $Y_{(1)}$ max is $Y_{(n)}$

We are interested in finding the distribution of an arbitrary $Y_{(i)}$ as well as the joint distributions of sets of $Y_{(i)}$ s and $Y_{(j)}$ s

ex: Range= $Y_{(n)} - Y_{(1)}$

7.3 r^{th} Order Statistic

We need to find the density of $Y_{(r)}$ at a value y

Consider 3 intervals $(-\infty, y)$, $[y, y + dy)$, $[y + dy, \infty)$

The number of observations in each of the intervals follows the tri-nomial distribution:

$$f(s_1, s_2, s_3) = \frac{n!}{s_1!s_2!s_3!} p_1^{s_1} p_2^{s_2} p_3^{s_3}$$

The event that $y \leq Y_{(r)} < y + dy$ is the event we have:

$(r - 1)$ observations are less than y ,

$(n - r)$ observations are greater than y

1 observation is in interval $y, y + dy$

In the trinomial distribution this corresponds to:

$$s_1 = r - 1, s_2 = 1, s_3 = n - r$$

$$p_1 = F_Y(y), p_2 = f_Y(y)dy, p_3 = 1 - F_Y(y + dy)$$

$$f_{Y_{(r)}}(y) = \frac{n!}{(r-1)!(n-r)!} F_Y(y)^{r-1} [1 - F_Y(y)]^{n-r} f_Y(y)$$

The above equation can also be written as:

$$\frac{F_Y(y)^{r-1} [1 - F_Y(y)]^{n-r} f_Y(y)}{B(r, n - r + 1)}$$

$$\text{Since } \frac{1}{B(n, n - r + 1)} = \frac{\Gamma(n + 1)}{\Gamma(r)\Gamma(n - r + 1)} = \frac{n!}{(r-1)!(n-r)!}$$

7.3.1 Example with $U(0,1)$

$F(y) = y$ that is $Y \sim U(0,1)$

$$f_{Y_{(r)}} = \frac{y^{r-1}(1-y)^{n-r}}{B(r, n - r + 1)}$$

$Y_{(r)}$ follows a beta($r, n - r + 1$) distribution

$$E[Y_{(r)}] = \frac{r}{n+1}$$

7.4 Distribution of the Median

Suppose sample size is odd: $n = 2m + 1$

Median is the $(m + 1)^{th}$ order statistic

set $r = m + 1$ $n = 2m + 1$ into rth order statistic formula giving us:

$$f_{med}(y) = f_{Y_{m+1}}(y) = \frac{F_Y(y)^m (1 - F_Y(y))^m f_Y(y)}{B(m+1, m+1)}$$

if the density $f_Y(y)$ is symmetric around zero, so that $EY = 0$ then:

$$F_Y(-y) = 1 - F_Y(y)$$

and so the density of the median is also symmetric around zero, so that:

$$E[med(Y_1, \dots, Y_n)] = 0$$

7.5 Joint Distribution of $X_{(i)}$ and $X_{(j)}$

Theorem 5.4.6 (from CB):

Let $X_{(1)}, \dots, X_{(n)}$ denote the order statistics of a random sample, X_1, \dots, X_n from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \leq i < j \leq n$, is:

$$f_{X_{(i)}, X_{(j)}}(u, v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) [F_X(u)]^{i-1} [F_X(v) - F_X(u)]^{j-1-i} [1 - F_X(v)]^{n-j}$$

for $-\infty < u, v < \infty$

7.5.1 Example with $U(0,1)$

$$F_Y(x) = x \quad U(0,1)$$

$$f_{Y_{(r)}, Y_{(s)}}(u, v) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} u^{r-1} (v-u)^{s-r-1} (1-v)^{n-s}$$

for $u < v$

7.6 Joint Distribution of all Order Statistics

The joint pdf of all the order statistics is:

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = \begin{cases} n! f_X(x_1) \cdots f_X(x_n) & -\infty < x_1 < \cdots < x_n < \infty \\ 0 & \text{otherwise} \end{cases}$$

7.7 Distribution of the Range

Set $r = 1$ and $s = n$ in theorem 5.4.6

This gives us the joint distribution of the min and max

$$f_{Y_{(1)}, Y_{(n)}}(u, v) = \frac{n!}{(n-2)!} [F_Y(v) - F_Y(u)]^{n-2} f_Y(u) f_Y(v)$$

Do a transformation to $R = Y_{(n)} - Y_{(1)}$ and $W = Y_{(1)}$
 The Jacobian is 1

$$f_{W,R}(w, r) = n(n-1)[F_Y(w+r) - F_Y(w)]^{n-2} f_Y(w) f_Y(w+r)$$

The density of R can be found by integrating out W:
 $f_R(r) = \int_{-\infty}^{\infty} f_{W,R}(w, r) dw$

7.7.1 Range Example

$$Y \sim U[0, 1] \quad F_Y(x) = x$$

$$\begin{aligned} f_R(r) &= \int_0^{1-r} n(n-1)r^{n-2} dw \\ &= n(n-1)r^{n-2}(1-r) \end{aligned}$$

R has a Beta distribution

$$\begin{aligned} E(R) &= n(n-1) \int_0^1 r * r^{n-2}(1-r) dr \\ &= n(n-1) \left[\frac{1}{n} - \frac{1}{n+1} \right] \\ &= \frac{n-1}{n+1} \end{aligned}$$