# Stats 512 ♡ 513 Review

Eileen Burns, FSA, MAAA

June 16, 2009

# Contents

# Chapter 1

# Basic Probability

## 1.1 Experiment, Sample Space, RV, and Probability

### Bell Curve (6)

Probability an observation falls within $x$ standard deviations of the mean of a normal random variable
$x = 1$: $P(-1 \leq Z \leq 1) = .682$
$x = 2$: $P(-2 \leq Z \leq 2) = .954$
$x = 3$: $P(-3 \leq Z \leq 3) = .9974$
Note: this is true for any $X \sim N(\mu, \sigma^2)$ with $Z = \dfrac{X-\mu}{\sigma}$.
From a quantile perspective,

| $p$ | .500 | .600 | .700 | .750 | .800 | .850 | .900 | .950 | .975 | .990 | .999 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $z_p$ | 0 | .254 | .524 | .672 | .840 | 1.37 | 1.28 | 1.645 | 1.96 | 2.33 | 3.80 |

### Basic probability facts (8)

$P(A^c) = 1 - P(A)$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

### Union-Intersection Principle (§1.7) (73)

$P(\cup_1^n A_i) = \sum_i P(A_i) = \sum_{i \neq j} P(A_i A_j) + \sum_{i \neq j \neq k} P(A_i A_j A_k) - \sum_{i \neq j \neq k \neq l} P(A_i A_j A_k A_l) + \cdots$

### Partition of Event (§1.7) (67)

For a given partition of the event $A$,
$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ when all pairs of events $A_i, A_j$ are disjoint.

### Density Functions and Distribution Functions (10)

For density function $f$ and distribution function $F$,
$P(a < X \leq b) = \int_a^b f(v)\, dv = \int_{-\infty}^b f(v)\, dv - \int_{-\infty}^a f(v)\, dv = F(b) - F(a)$

## 1.3 Expected Value Descriptive Parameters

| | **Discrete** | **Continuous** |
|---|---|---|
| **Expected Value** | (26) | (31) |
| $\mu_X = E[X]$ | $\sum_{all v} v \cdot p_X(v)$ | $\int_{-\infty}^{\infty} v f(v)\,dv$ |
| $E[g(X)]$ | $\sum_{all v} g(v) \cdot p_X(v)$ | $\int_{-\infty}^{\infty} g(v) f(v)\,dv$ |
| **Mean deviation** | (27) | (31) |
| $\tau_X = E[\lvert X - \mu\rvert]$ | $\sum_{all v} \lvert v - \mu_X\rvert \cdot p_X(v)$ | $\int_{-\infty}^{\infty} \lvert v - \mu_X\rvert f(v)\,dv$ |
| **Variance** | | |
| $\sigma_X^2 = E[X^2] - (E[X])^2$ | $\sum_{all v} (v - \mu_X)^2 \cdot p_X(v)$ | $\int_{-\infty}^{\infty} (v - \mu_X)^2 f(v)\,dv$ |

## 1.4 More About Normal Distributions

## 1.5 Independent Trials and a Pictorial CLT

### Central Limit Theorem (46)

Let $X_i$ for $1 \le i \le n$ denote iid rvs.
Let $T_n = X_1 + X_2 + \cdots + X_n$.
Let $Z$ denote a $N(0,1)$ rv.
Then for all $-\infty \le a < b \le \infty$, $P\left(a \le \frac{T_n - n\mu x}{\sqrt{n}\sigma_x} \le b\right) \to P(a \le Z \le b)$ as $n \to \infty$.

## 1.6 The Population, the Sample, and Data

## 1.7 Elementary Probability, Stressing Independence

### Conditional Probability (58)

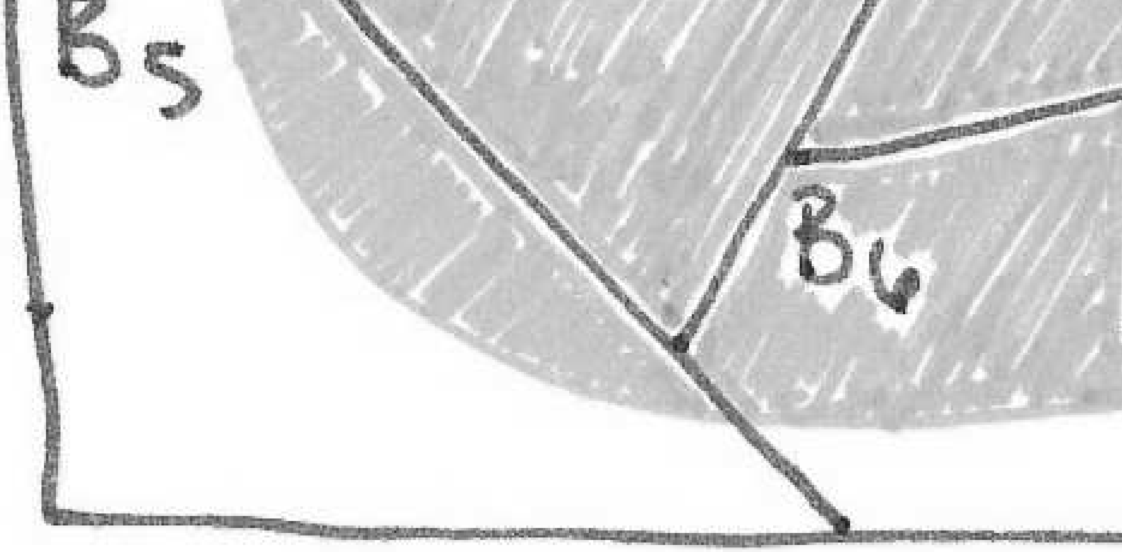
$P(A|B) = \frac{P(A \cap B)}{P(B)}$ $\qquad$ $P(A \cap B) = P(A|B) \cdot P(B)$

### Independent Events (62)

$P(A|B) = P(A)$ $\qquad$ $P(A \cap B) = P(A) \cdot P(B)$

### General Independence of Random Variables (64)

1. $X_1$ and $X_2$ are independent rvs
2. $P(X_1 \in A \text{ and } X_2 \in B) = P(X_1 \in A)P(X_2 \in B)$ for all (one-dimensional) events $A$ and $B$.




### Bayes Theorem (67)

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A|B_i) \cdot P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)}$$

### Marginals (67)

$p_X(v) = P(X = v) = \sum_{allw} P(X = v \text{ and } Y = w) = \sum_{allw} p_{X,Y}(v, w).$

### Example

Maximums and minimums of independent rvs (74)

## 1.8 Expectation, Variance, and the CLT

### Sums of Arbitrary RVs (83)

$\mu_{aX+b} = a\mu_X + b$ $\quad \mu_{aX+bY} = a\mu_X + b\mu_Y$ $\quad \sigma^2_{aX+b} = a^2\sigma^2_X$

### Sums of Independent RVs (83)

$E(XY) = E(X) \cdot E(Y) = \mu_X \cdot \mu_Y$ $\quad Var(X + Y) = Var(X) + Var(Y) = \sigma^2_X + \sigma^2_Y$

### Mean and Variance of a Sum (83)

For independent repetitions of a basic $X$-experiment $X_1, \ldots, X_n$, Let $T_n = X_1 + \cdots X_n$. Then
$E(T_n) = n\mu_x$
$Var(T_n) = n\sigma^2_x$

$Z_n = \dfrac{T_n - n\mu_X}{\sqrt{n}\sigma_X} = \dfrac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}}$ has mean 0 and variance 1.

### Covariance (86)

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

### Correlation (90)

Corr$[X, Y]$ =Cov$[X, Y]/\sigma_X\sigma_Y$

### Sample Moments (87)

$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

$\hat{\sigma}_{X,Y} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$

### Handy formulas (89)

$\sigma^2 = E[X(X-1)] + \mu - \mu^2$
$\sigma^2 = E[X(X+1)] - \mu - \mu^2$

## 1.9 Applications of the CLT

Continuity correction (93)
Pretty graphs (99)

# Chapter 2

# Introduction to Statistics

## 2.1 Presentation of Data

See Appendix B for examples of plot types.

## 2.2 Estimation of $\mu$ and $\sigma^2$

| | Estimator | Expected Value | Variance | |
|---|---|---|---|---|
| Sample mean | $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ | $E(\bar{X}_n) = \mu$ | $\mathrm{Var}[\bar{X}_n] = \sigma^2/n$ | (114) |
| Sample variance | $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ | $E[S_n^2] = \sigma^2$ | $\mathrm{Var}[S_n^2] = \frac{\mu_4}{n} - \frac{1}{n}\frac{n-3}{n-1}\sigma^4$ | (115) |

## 2.3 Elementary Classical Statistics

### Actual accuracy ratio of $\bar{X}_n(T_n)$ (121)

$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \cong N(0, 1)$

### CLT of Probability (121)

No matter what the shape of the distribution of the population of possible outcomes of an original $X$-experiment that has mean $\mu$ and standard deviation $\sigma$, the following result will hold true. As $n$ gets larger, the distribution of the standardized $Z_n$ form of the $\bar{X}_n$-experiment gets closer to the distribution given by the standardized bell shaped Normal curve.

### Estimated accuracy ratio of $\bar{X}_n(T_n)$ (122)

$T_n = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \cong T_{n-1}$

### CLT of Statistics (122)

The distribution of the ratio $T_n$ is given (not by the standardized bell curve, but rather) by the Student $\mathcal{T}_{n-1}$ density. This is an approximation that gets better as the sample size $n$ gets larger.

### Random confidence intervals vs. hypothesis testing vs $p$-values (123)

**Random confidence intervals:** 95% CI for $\mu$ consists of exactly those values of $\mu$ that are .95-plausible
**Hypothesis testing:** .05-level test of $H_o : \mu = \mu_o$ rejects $H_o$ at the .05 level if $\mu_o$ is not .95-plausible
**p-values:** 2-sided p-value associated with $\mu_o$ is just exactly that level at which $\mu_o$ first becomes plausible

### Theorem 2.3.1 – Student $T$-statistic (124)

If $X_1, X_2, \ldots, X_n$ are independent rvs that all have *exactly* a $N(\mu, \sigma^2)$ distribution, then Students estimated accuracy ratio has *exactly* the $\mathcal{T}_{n-1}$ distribution.

### Metatheorem 2.3.1 (126)

If $X_1, X_2, \ldots, X_n$ are independent rvs that all have *approximately* a $N(\mu, \sigma^2)$ distribution, then Students estimated accuracy ratio has *approximately* the $\mathcal{T}_{n-1}$ distribution.

### Metatheorem 2.3.2 (126)

If $X_1, X_2, \ldots, X_n$ are *nearly* independent rvs that have distributions that are not too skewed and do not produce severe outliers, then the following estimated accuracy ratio is *approximately* Normal, or some $\mathcal{T}$: $\dfrac{\bar{W} - \mu}{S}$.

### Student $\mathcal{T}_n$ density (128)

$f_n(t) = \dfrac{\Gamma((n+1)/2)}{\sqrt{\pi n}\,\Gamma(n/2)} \dfrac{1}{(1 + t^2/n)^{(n+1)/2}} \to \dfrac{1}{\sqrt{2\pi}} e^{-t^2/2}$.

## 2.4 Elementary Statistical Applications

### Type I error (130)

Rejecting the null hypothesis $H_o$ when true

### Type II error (130)

Not rejecting $H_o$ when it is false

### Test statistic $T_n$ (130)

$T_n(\mu_o) = \frac{\bar{X}_n - \mu_o}{s_n/\sqrt{n}} \cong T_{n-1}$

# Chapter 3

# Probability Models

## 3.1 Math Facts

### Geometric Series (135)

$$S_n = \sum_{k=0}^{n} ar^k = \frac{a}{1-r}(1 - r^{n+1}) \qquad S_\infty = \sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$$

### Infinite Series (135)

$$f(x) = \sum_{k=0}^{\infty} a_k x^k \qquad f'(x) = \sum_{k=0}^{\infty} k a_k x^{k-1}$$

### Exponential and Logarithmic Series (135)

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \qquad log(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1} x^k}{k} \text{ for } |x| < 1$$

### Fundamental Exponential Limit (136)

$\left(1 + \frac{a_n}{n}\right)^n \to e^a$ as $n \to \infty$, whenever $a_n \to a$ as $n \to \infty$.

### Mean Value Theorem (136)

$f(b) - f(a) = (b-a)f'(c)$ for some $a < c < b$.

### L'Hospital's Rule (136)

Suppose $f(a) = g(a) = 0$, and that $f'(a)$ and $g'(a) \neq 0$ exist. Then
$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$.

### Taylor Series Expansion (136)

$f(x) = f(a) + \frac{(x-a)}{1!} f'(a) + \cdots + \frac{(x-a)^k}{k!} f^{(k)}(a) + R_k(x)$

### Integration by Parts (137)

$\int u dv = uv - \int v du$

**Gamma Function & Properties** **(138)**

$\Gamma(r) = \int_0^\infty y^{r-1}e^{-1}dy$ $\quad\Gamma(r) = (r-1)\Gamma(r-1)$ $\quad\Gamma(r) = (r-1)!\Gamma(\frac{1}{2}) = \sqrt{\pi}$

**Standard Normal Density Function & Moments** **(139)**

$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ $\quad\mu_x = 0$ $\quad\sigma_x^2 = 1$

**Gamma Density & Moments** **(140)**

$f(x) = \frac{1}{\Gamma(r)}\frac{w^{r-1}}{\theta^r}e^{-w/\theta}$ $\quad\mu_w = r\theta$ $\quad\sigma_w^2 = r\theta^2$

## 3.2 Combinatorics and Hypergeometric RVs

**Permutations & Combinations** **(148)**

$P_n^k = \frac{n!}{(n-k)!}$ $\quad C_n^k = \binom{n}{k} = \frac{n!}{(n-k)!k!}$

**Sampling with replacement (Binomial)** **(150)**

**Sampling without replacement (Hypergeometric$(R, W, n)$)** **(151)**

$P(X = k) = \frac{\binom{R}{k}\binom{W}{n-k}}{\binom{N}{n}}$ $\quad\mu_x = \frac{nR}{N}$ $\quad\sigma_x^2 = \frac{nR}{N} \cdot \frac{(N-R)(N-n)}{N(N-1)}$

**Mathematics via revisualization** **(157)**

$\binom{N}{k} = \binom{N}{N-k}$ $\quad\binom{N}{k} = \binom{N-1}{k} + \binom{N-1}{k-1}$ $\quad\binom{m}{n} = \binom{n-1}{n-1} + \binom{n}{n-1} + \binom{n+1}{n-1} + \binom{n+2}{n-1} + \cdots + \binom{m-1}{n-1}$

## 3.3 Independent Bernoulli Trials

**Bernoulli$(p)$ − $X_i$ = independent trials** **(171)**

$P(X = 1) = p$ $\quad\mu_{x_i} = p$ $\quad\sigma_{x_i}^2 = qp$
$P(X = 0) = 1 - p = q$

**Binomial$(n, p)$ − $T_n$ = sum of independent trials** **(173)**

$P(T_n = k) = \binom{n}{k}p^k q^{n-k}$ $\quad\mu_{T_n} = np$ $\quad\sigma_{T_n}^2 = npq$

**Geometric Turns − GeoT$(p)$ − $Y_i$ = interarrival times** **(174)**

$P(Y = k) = q^{k-1}p$ $\quad\mu_{Y_i} = 1/p$ $\quad\sigma_{Y_i}^2 = q/p^2$
$P(Y > k) = q^k$

**GeoT$(p)$ lack of memory property** **(174)**

$P(Y > i + k | Y > k) = P(Y > i) = q^i$

**Negative Binomial Turns − NegBiT$(r, p)$ − $W_r$ = waiting times** **(176)**

$P(W_r = k) = \binom{k-1}{r-1}p^r q^{k-r}$ $\quad\mu_{W_r} = r/p$ $\quad\sigma_{W_r}^2 = rq/p^2$

**Event Identity** **(176)**

$[W_r > n] = [T_n < r]$
$[W_r = k] = [T_{K-1} = r-1] \cap [X_k = 1]$

**Sums of NegBiT rvs** **(177)**

NegBiT$(r,p)$+NegBiT$(s,p)$ =NegBiT$(r+s,p)$

**Geometric Failures − GeoF$(p)$ − $Y'$ = failures before first success** **(177)**

$P(Y'=k) = q^k p$ $\quad Y'+1 = Y$ $\quad \mu_{Y'} = q/p$ $\quad \sigma^2_{Y'} = q/p^2$

**Negative Binomial Failures − NegBiF$(r,p)$ − $W'_r$ = failures before the $k^{th}$ success (177)**

$P(W'_r = k) = \binom{k+r-1}{r-1} p^r q^k$ $\quad W'_r + r = W_r$ $\quad \mu_{W'_r} = rq/p$ $\quad \sigma^2_{W'_r} = rq/p^2$

**Examples**

Comparing two success probabilities $p_1$ and $p_2$ (185)
Gamber's Ruin (194)

## 3.4 The Poisson Distribution

**Poisson$(\lambda)$ − $\mathbb{N}_t$ = count by time t** **(208, 219)**

$P(T=k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$ $\quad \mu_T = \lambda$ $\quad \sigma^2_T = \lambda$

**Poisson Limit Theorem** **(208)**

$T_n$ ∼Binomial$(n, p_n)$ $\quad \lambda_n = np_n \to \lambda$ as $n \to \infty$ $\quad P(T_n = k) \to P(T=k) = \frac{\lambda^k}{k!} e^{-\lambda}$

**Sums of Poissons is Poisson** **(210)**

## 3.5 The Poisson Process $\mathbb{N}$

**Poisson$(\nu t)$ − $\mathbb{N}_t$ = count by time t** **(208, 219)**

$P(T=k) = \dfrac{(\nu t)^k}{k!} e^{-\nu t}$ $\quad \mu_T = \nu t = \lambda$ $\quad \sigma^2_T = \nu t = \lambda$

**Exponential $(\theta)$ − $Y_i$ = interarrival times** **(220)**

$f_Y(t) = \nu e^{-\nu t} = \frac{1}{\theta} e^{-t/\theta}$ $\quad \mu_{Y_i} = 1/\nu = \theta$ $\quad \sigma^2_{Y_i} = (1/\nu)^2 = \theta^2$

**Event Identity** **(220)**

$[Y > t] = [N_t = 0] = [N_t < 1]$

**Exponential $(\theta)$ lack of memory** **(220)**

$P(Y > s+t | Y > s) = P(Y > t) = e^{-\nu t} = e^{-t/\theta}$

### Gamma($r$, "$rate$" $= \nu$) $- W_r =$ waiting times (221)

$f_{W_r}(t) = \frac{1}{\Gamma(r)}\nu^r t^{r-1} e^{-\nu t} = \frac{1}{\Gamma(r)}\frac{t^{r-1}}{\theta^r} e^{-t/\theta}$ $\mu_{W_r} = r/\nu = r\theta$ $\sigma^2_{W_r} = r/\nu^2 = r\theta^2$

### Event Identity (221)

$[W_r > t] = [N(t) < r]$

### Sums of Gamma RVs (222)

Gamma$(r,\nu)$ + Gamma$(s,\nu)$ = Gamma$(r+s,\nu)$

### Random location of the times of event occurrences is Binomial (222, 214)

$P(T_1 = k|T_1 + T_2 = n) = \binom{n}{k} p^k (1-p)^{n-k}$ where $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$

### Conditioned Poisson sums have Binomial Distributions (222, 214)

$(N_s|N_t = m)$ ~Binomial$(n, p = \frac{s}{t})$

### Example

Comparison of bulb lifetimes (230)

## 3.6 The Failure Rate Function $\lambda(\cdot)$

### Failure Rate Function (241)

With a constant failure rate, $\lambda(t) = \frac{f(t)}{1-F(t)} = F'(t)/[1-F(t)] = \lambda$ for all $t \geq 0$.
For a nonconstant failure rate, this generalizes to $\frac{d}{dt} log[1 - F(t)] = -\lambda(t)$.

## 3.7 Min, Max, Median, and Order Statistics

### Uniform(0,1) order statistics (252)

Let $U_1, \ldots, U_n$ be idnependent Uniform(0,1) rvs.
Let $0 < U_{n:1} < \cdots < U_{n:n} < 1$ denote their order statistics.
Then $B_n(t) = \sum_{k=1}^n I_{[0,t]}(U_k)$ = (the number of $U_1, \ldots, U_n$ that are $\leq t$) is a Binomial$(n,t)$ rv.

### Event Identity (252)

$[U_{n:k} > t] = [B_n(t) < k]$

### Distribution of Uniform Order Statistics (252)

$U_{n:k}$ has the Beta$(k, n-k+1)$ density.
$f_{U_{n:k}}(t) = \binom{n}{k-1,1,n-k} t^{k-1}(1-t)^{n-k}$ for $0 \leq t \leq 1$.

### Asymptotic Distribution of the minimum of Uniform(0,1) rvs (258)

$P(nU_{n:1} > t) = P(X > t/n)^n = (1 - t/n)^n \to e^{-t}$ for all $0 \leq t < n$.

### Joint Distribution of Uniform Order Statistics (253)

$f_{U_{n:i},U_{n:k}}(s,t) = \binom{n}{i-1,1,k-i-1,1,n-k} s^{i-1}(t-s)^{k-i-1}(1-t)^{n-k}$ for $0 \leq s < t \leq 1$
$Cov[U_{n:i}, U_{n:k}] = \frac{1}{n+2}\frac{i}{n+1}\big(1 - \frac{k}{n+1}\big)$ for all $1 \leq i < k \leq n$

### Distribution of General Order Statistics (253)

$f_{X_{n:k}}(x) = \binom{n}{k-1,1,n-k} F(x)^{k-1} f(x)[1-F(x)]^{n-k}$ for all $x$.

### Joint Distribution of General Order Statistics (253)

$f_{X_{n:i},X_{n:k}}(x,y) = \binom{n}{i-1,1,k-i-1,1,n-k} F(x)^{i-1} f(x)[F(y)-F(x)]^{k-i-1} f(y)[1-F(y)]^{n-k}$ for $x \leq y$

### Example

Exponential Order Statistics (256)

## 3.8 Multinomial Distributions

### Multinomial($n; \mathsf{p}$) Distribution of $\mathsf{N}$ (263)

$P(\mathsf{N} = \mathsf{m}) = \binom{n}{m_1, m_2, \ldots, m_I} p_1^{m_1} p_2^{m_2} \cdots p_I^{m_I}$ for each possible outcome $\mathsf{m} \geq 0$ having $\sum_1^I m_i = n$.

### Marginal Distributions (264)

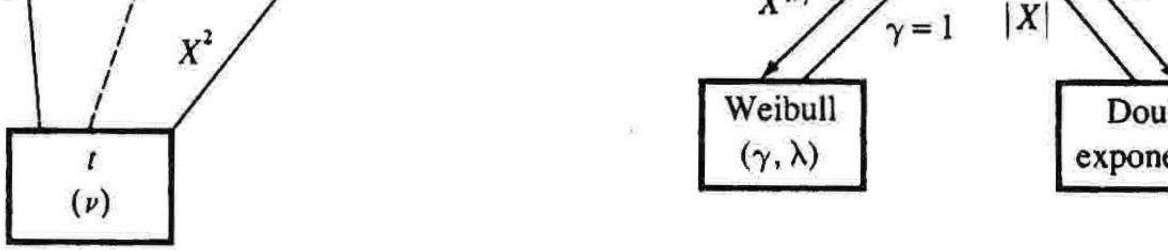
$N_i \sim$ Binomial$(n, p_i)$.

| | | | | |
|---|---|---|---|---|
| $\mathrm{Cov}[X_{ki}, X_{\ell j}]$ | $=$ | $p_i(1-p_i)$ for $i = j$ | $=$ | $-p_i p_j$ for $i \neq j$ |
| $\mathrm{Cov}[N_i, N_j]$ | $=$ | $np_i(1-p_i)$ | $=$ | $-np_i p_j$ |
| $\mathrm{Cov}[\hat{p}^i, \hat{p}^j]$ | $=$ | $\frac{1}{n} p_i(1-p_i)$ | $=$ | $-\frac{1}{n} np_i p_j$ |

### Conditional Distributions (265)

Given that $T = N_{I_1+1} + \cdots + N_I = t$
$((N_1, \ldots, N_{I_1}) | T = t) \sim$Multinomial$(n-t; p'_1, \ldots, p'_{I_1})$

## 3.9 Sampling from a Finite Populations, with a CLT




**Relationships among common distributions.** Solid lines represent transf
special cases, dashed lines represent limits. Adapted from Leemis (1986).

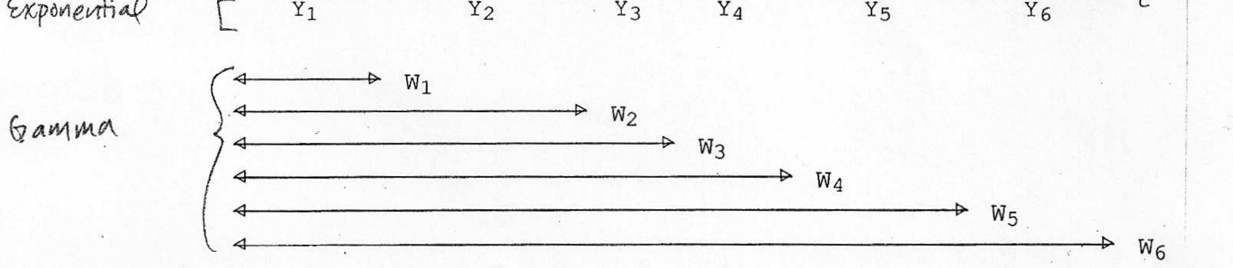



Figure 3.5.1 RVs associated with the Poisson Process

# Chapter 4

# Dependent Random Variables

## 4.1 Two-Dimensional Discrete RVs

**Joint Probability Distribution** **(288)**

$P_{X,Y}(x,y) = P(X = x, Y = y)$

**Marginal distributions of X and Y** **(288)**

$P_X(x) = \sum_{\forall y} P_{X,Y}(x,y)$ $P_Y(y) = \sum_{\forall x} P_{X,Y}(x,y)$

**Independence** **(288)**

$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$ $P_{X,Y}(v,w) = P_X(v) \cdot P_Y(w) \quad \forall v, w$

**Convolution formula (for $Z = X + Y$)** **(290)**

$$\begin{aligned} P_Z(k) = P(Z = k) = P(X + Y = k) &= \sum_{i=-\infty}^{\infty} P(X = i, Y = k - i) \\ &= \sum_{i=-\infty}^{\infty} P_X(i) \cdot P_Y(k - i) \\ &= \sum_{i=0}^{k} P_X(i) \cdot P_Y(k - i) \end{aligned}$$

**Conditional distribution of $Y$ given $X = x$** **(292)**

$$p_{Y|X=x}(y) = P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

$p_{X,Y}(x,y) = p_X(x) \cdot p_{Y|X=x}(y) = p_Y(y) \cdot p_{X|Y=y}(x)$

**Expected value of two-dimensional discrete RVs** **(294)**

$$E[g(X,Y)] = \sum_{\forall v} \sum \forall w g(v,w) \cdot P((X,Y) = (v,w))$$

$$E[g(X)] = \sum_{\forall v} g(v) \sum_{\forall w} P_{X,Y}(v,w) = \sum_{\forall v} g(v) P_X(v)$$

**Theorem of the Unconscious Statistician** **(295)**

If $V = g(X,Y)$ then $E(V) = \sum_{\forall v} v p_V(v) = E[g(X,Y)] = \sum_{\forall x} \sum_{\forall y} g(x,y) p(x,y)$.

**Independence of Partitions** **(297)**

$$P(A_i, B_j) = p_{X,Y}(a_i, b_j) = p_X(a_i) \cdot p_{Y|X=x}(b_j) = P(A_i) \cdot P(B_j|A_i) \quad \forall i, j$$
$$= p_X(a_i) \cdot p_Y(b_j) = P(A_i) \cdot P(B_j)$$

## 4.2 Two-Dimensional Continuous RVs

**Joint density function** **(304)**

$$P((X,Y) \in A) = \iint\limits_A f(x,y)\,dx\,dy$$

**Joint distribution function** **(304)**

$$F(x,y) = F_{X,Y}(x,y) = P(X \le x, Y \le y)$$

**Marginal densities** **(304)**

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

**Conditional densities** **(305)**

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} \Rightarrow f(x,y) = f_X(x) \cdot f_{Y|X=x}(y)$$
$$\Rightarrow f(x,y) = f_X(x) \cdot f_Y(y)$$

**Expected value of two-dimensional continuous RVs** **(305)**

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\,dx\,dy$$

**Convolution formula for the sum of continuous RVs** $(Z = X + Y)$ **(306)**

$$\begin{aligned}
F_Z(z) = P(Z \le z) = P(X+Y \le z) &= \iint_{[u+v \le z]} f(u,v)\,du\,dv \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-v} f(u,v)\,du\,dv && \text{for any joint density } f \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(u, w-u)\,dw\,du && \text{letting } w = v + u \\
&= \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(u, w-u)\,du\,dw && \text{always ok when } f \ge 0 \\
&= \int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f_X(u) f_Y(w-u)\,du\right] dw && \text{when } X \text{ and } Y \text{ are independent.}
\end{aligned}$$

**Theorem of the Unconscious Statistician** **(307)**

If $Z = g(X,Y)$, then $E(Z) = \int_{-\infty}^{\infty} z f_Z(z)\,dz = E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\,dx\,dy.$

**Sums of Independent rvs** **(308)**

Sums of independent Normals are Normal
Sums of independent Cauchy rvs are Cauchy
Sums of independent Gammas are Gamma

Sums of two independent Uniform rvs have a Triangular density (see Appendix C)

### Factoring (74,309)

Random variables are independent if and only if their distribution factors

### Example: Density of a quotient (309)

The quotient $R = U/V$ has density $f_R(r) = \int_{-\infty}^{\infty} v f(rv, v) dv$ for all $r$.

## 4.3 Conditional Expectation

### Conditional probability distribution of $Y$ given $X = x$ (312)

$$p_{Y|X=x}(y) = P(Y = y|X = x) = \frac{p(x, y)}{p_X(x)}$$

### Conditional mean (regression function) (312,328)

$$m(x) = E(Y|X = x) = \sum_{\forall y} y\, p_{Y|X=x}(y)$$

$$\mu_Y = E(Y) = E[E(Y|X = x)] = E[m(x)]$$

### Conditional variance (312)

$$v^2(x) = Var(Y|X = x) = \sum_{\forall y} (y - m(x))^2\, p_{Y|X=x}(y)$$

$$\sigma_Y^2 = Var(Y) = E[Var(Y|X = x)] + Var[E(Y|X = x)] = E[v(x)] + Var[m(x)]$$

### Conditional expectation of $g(X, Y)$ given $X = x$ (312)

$$E[g(X, Y)|X = x] = \sum_{\forall y} g(x, y)\, p_{Y|X=x}(y)$$

$$E[g(X, Y)] = E[E[g(X, Y)|X = x]]$$

### A Factoring Example (313)

For $0 \le y \le x \le 1$, $f(x, y) = \frac{1}{x}$

$$= \frac{1}{x} \cdot 1_{[0 \le y \le x \le 1]}$$

$$= 1_{[0,1]}(x) \cdot [\tfrac{1}{x} \cdot 1_{[0,x]}(y)]$$

$$= f_X(x) \cdot f_{Y|X=x}(y)$$

### Summing a random number of RVs (315)

Let $X_1, X_2, \ldots, X_N$ be iid with mean $\mu_X$ and $N \ge 0$.

Then for $T = \sum_{i=1}^{N} X_i$, $\quad E(T) = \mu_X \mu_N \quad Var(T) = \mu_N \sigma_X^2 + \mu_X^2 \sigma_N^2$

### Example

Conditioning via regeneration (317)

## 4.4 Prediction

**Mean squared error** **(326)**

$E[(Y-a)^2]$

**Least squares predictor of $Y = \mu_F$** **(326)**

$\bar{a}_F$ such that $E[(Y-\bar{a}_F)^2] = min_a E[(Y-a)^2]$

**Mean absolute deviation** **(326)**

$E[|Y-a|]$

**Least absolute deviation predictor of $\ddot{Y}(\ddot{a}_F)$** **(326)**

$E[|Y-\ddot{a}_F|] = min_a E[|Y-a|]$

**Bias = systematic error** **(326)**

$b_F = a_F - \mu_F =$ difference between predictor and mean
$E[(Y-a_F)^2] = \sigma_F^2 + b_F^2 =$ mean squared error is the variance plus bias squared

**Median ($\ddot{m}_F$)** **(327)**

$$\int_{-\infty}^{\ddot{m}_F} f_Y(y)\,dy = 0.5$$

**Analogs in Samples** **(327)**

$\hat{L}S$ predictor of a future observation $Y = \bar{Y}$
$\hat{L}AD$ predictor of a future observation $Y = \ddot{Y}$=median$(Y_1, \ldots, Y_n)$

**Other moments** **(329)**

$E[X^k]$ $k^{th}$ moment
$E[(X-\mu)^k]$ $k^{th}$ central moment
$E[|X|^r]$ $r_{th}$ absolute moment
$E[|X-\ddot{\mu}|]$ mean deviation about the median ($\ddot{\mu}$ is $a$ median of $X$)

**Least squares predictor** **(330)**

**Best linear predictor** **(330)**

## 4.5 Covariance

**Covariance** **(333)**

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

**Correlation** **(333)**

$$\rho_{X,Y} = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = Cov\left[\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right]$$

**Correlation inequality** **(335)**

$-1 \leq \rho \leq 1$

$\rho = \pm 1$ iff $(Y-\mu_y) = \pm\frac{\sigma_Y}{\sigma_X}(X-\mu_X)$

**Sample variance, population variance** **(336)**

$$s^2 = s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n-1}SS_{XX} \qquad \hat{\sigma}^2 = \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 = \frac{1}{n}SS_{XX}$$

$E[S^2] = \sigma^2$

$$Var[S^2] = \frac{\mu_4}{n} - \frac{1}{n}\frac{n-3}{n-1}\sigma^4 = \frac{\mu_4 - \sigma^4}{n} + \frac{2\sigma^4}{n(n-1)} \approx \frac{\mu_4-\sigma^4}{n}$$

**Sample covariance** **(336)**

$s_{X,Y} = \frac{1}{n-1}SS_{XY}$ $\quad$ $\hat{\sigma}_{X,Y} = \frac{1}{n}SS_{XY} = \overline{XY} - \bar{X}\bar{Y}$ $\quad$ where $SS_{XY} = \sum(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$

**Sampe correlation coefficient, population correlation coefficient** **(336)**

$$\hat{\rho} = \hat{\rho}_{X,Y} = \frac{s_{X,Y}}{s_X s_Y} = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X \hat{\sigma}_Y} = \frac{SS_{XY}}{\sqrt{SS_{XX}SS_{YY}}}$$

**Empirical distribution (distribution of the sample)** **(337)**

$E(U) = \bar{X}$ $\quad$ $E(V) = \bar{Y}$ $\quad$ $Cov(U,V) = \hat{\sigma}_{X,Y}$

$Var(U) = \hat{\sigma}_X^2$ $\quad$ $Var(V) = \hat{\sigma}_Y^2$ $\quad$ $Corr(U,V) = \hat{\rho}_{X,Y}$

### Covariance of linear combinations (337)

$$\begin{aligned} Cov[X+Y, Y+Z] &= Cov[X,Y] + Cov[X,Z] + Cov[Y,Y] + Cov[Y,Z] \\ &= Cov[X,Y] + Cov[X,Z] + Var(Y) + Cov[Y,Z] \end{aligned}$$

### Covariance of general linear combinations (338)

### Union/intersection formula (339)

### Poisson Limit Theorem (redux) (339)

### One-sample Covariance structure (341)

$$\begin{aligned} Cov[X_i, \bar{X}] &= \tfrac{1}{n}\sigma^2 \\ Var[X_i - \bar{X}] &= \tfrac{n-1}{n}\sigma^2 \\ Cov[X_i - \bar{X}, X_j - \bar{X}] &= -\tfrac{1}{n}\sigma^2 \\ Corr[X_i - \bar{X}, X_j - \bar{X}] &= -\tfrac{1}{n-1} \end{aligned}$$

## 4.6 Bivariate Normal Distributions

### Bivariate Normal Distribution (348)

$X$ and $Y$ have a Bivariate Normal Distribution:

$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} x_\mu \\ y_\mu \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \sigma_{X,Y} \\ \sigma_{X,Y} & \sigma_Y^2 \end{bmatrix}\right)$, where $\sigma_{X,Y} = \rho\sigma_X\sigma_Y$

with

$f(x,y) = \dfrac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} exp(-\dfrac{1}{2(1-\rho^2)}Q(x,y))$

where $Q(x,y) = \left(\dfrac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\dfrac{x-\mu_X}{\sigma_X}\right)\left(\dfrac{y-\mu_Y}{\sigma_Y}\right) + \left(\dfrac{y-\mu_Y}{\sigma_Y}\right)^2$

### Marginal distribution, Conditional distribution (348)

$f(x,y) = f_X(x) f_{Y|X=x}$, where

$f_X(x) \sim N(\mu_X, \sigma_X^2)$

$f_{Y|X=x}(y) = \dfrac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} exp\left(-\frac{1}{2}\left(\dfrac{y - [\mu_Y + (\rho\frac{\sigma_Y}{\sigma_X})(x-\mu_X)]}{\sigma_Y\sqrt{1-\rho^2}}\right)^2\right) \sim N(m(x), v^2(x))$

### Regression function & parameters (349)

$$\begin{aligned} m(x) = E[Y|X=x] &= \mu_Y + (\rho\tfrac{\sigma_Y}{\sigma_X})(x-\mu_X) \\ &= \alpha + \beta(x-\mu_X) && \text{where } \alpha = \mu_Y, \quad \beta = \rho\tfrac{\sigma_Y}{\sigma_X} \\ &= \gamma + \beta x && \text{where } \gamma = \mu_Y - \beta\mu x \end{aligned}$$

$$v^2(x) = Var(Y|X=x) = \sigma_Y^2(1-\rho^2) = \sigma_\epsilon^2 \qquad \text{where } \sigma_\epsilon = \sigma_Y\sqrt{1-\rho^2}$$

### Independence (349)

Bivariate Normal rvs $(X,Y)$ are independent if and only if $\rho = 0$.

## Factorization (352)

Any factorization of the type $g(x)h_x(y))$ (in which all $h_x(y)$ are dentsities) is a *unique* factorization, in general.

## Regression line $(m(x))$ (355)

$y = \rho x$ for standard normals
$y = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ for general normals

## Standard deviation line/line of symmetry for contour ellipses (355)

$y = x$ for standard normals
$y = \mu_Y + \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ for general normals

## Extra Facts (356)

Linear functions of Bivariate Normals are Bivariate Normal (see section 5.2)
Elliptical contours (357)
Contour lines (orthogonal transformations) (357)

# Chapter 5

# Distribution Theory

## 5.1 Transformations of RVs

### Linear transformations (365)

$Y = aX + b$

### Distribution function method (365)

Note that:
$F_Y(y) = P(Y \le y) = P(aX + b \le y) = P(X \le (y-b)/a) = F_X((y-b)/a)$

Differentiating both sides yields:
$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X((y-b)/a) = \frac{1}{a} f_X((y-b)/a)$

### Linear transformations of Normal RVs (365)

$$\begin{array}{llllll}
Z & \sim & N(0,1) & & \\
X = \mu + \sigma Z & \sim & N(\mu, \sigma^2) & & \\
Y = aX + b & \sim & (a\mu + b) + (a\sigma)Z & \sim & N(a\mu + b, a^2\sigma^2) \\
Z^2 = (N(0,1))^2 & \sim & Gamma(\frac{1}{2}, \text{"}mean\text{"} = 2) & \sim & \chi^2_1 \\
\sum Z_i^2 & \sim & \sum(\text{Gamma}(1/2, \text{"mean"} = 2))^2 & \sim & \chi^2_{2r}
\end{array}$$

### Scale transformations of Gamma RVs (365)

$$\begin{array}{lllll}
Y & \sim & \text{Gamma}(r, \theta) & & \\
Y/\theta & \sim & \text{Gamma}(r, \text{"mean"} = 1) & & \\
2Y/\theta & \sim & \text{Gamma}(r, \text{"mean"} = 2) & \sim & \chi^2_{2r}
\end{array}$$

### Distribution function method for monotone transformations (365)

For *monotone transformations* (strictly increasing & decreasing functions $Y = g(X)$),
$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))$

Therefore,
$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X((y-b)/a) = \frac{1}{a} f_X((y-b)/a)$.

Stated differently,
$f_Y(y) = f_X(x)\frac{dx}{dy} = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) = f_X(g^{-1}(y)) \cdot \frac{1}{g'(g^{-1}(y))}$

### Change of variable formula (365)

$f_Y(y) = f_X(x) \cdot |\frac{dx}{dy}| = f_X(x) \cdot \frac{1}{|dy/dx|}$ OR $f_{New}(y) = f_{Old}(x) \cdot |\frac{dOld}{dNew}|$

### Distribution function method for multiple valued transformation (366)

For $Y = X^2$,
$F_Y(y) = P(Y \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y} - 0)$

$f_Y(y) = [f_X(\sqrt{y}) + f_X(-\sqrt{y})]/2\sqrt{y}$

### Probability integral transformation (366)

$U = F(X) \sim U(0,1)$
$F_U(u) = P(U \le u) = P(F(X) \le u) = P(X \le F^{-1}(u)) = F(F^{-1}(u)) = F(u)$

### Inverse transformation (367)

$U \sim U(0,1); X = F^{-1}(u)$
$F_X(x) = P(X \le x) = P(F^{-1}(u) \le x) = P(u \le F(x)) = F(x)$

### Transformation by integration (367)

$F_W(x) = P(W \le x) = \int \cdots \int_{[x:H(x) \le w]} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$

### Student-t, F, Beta (369)

For $Z \sim N(0,1), \quad U \sim \chi^2_m, \quad V \sim \chi^2_n$ (independent),

$$\begin{aligned} T &= \frac{Z}{\sqrt{V/n}} \sim \mathcal{T}_n \\ W &= \frac{\frac{1}{m}U}{\frac{1}{n}V} \sim \mathcal{F}_{m,n} \\ B &= \frac{U}{U+V} \sim \text{Beta}(\tfrac{m}{2}, \tfrac{n}{2}) \end{aligned}$$

### Non-central $T$-distribution, $\chi^2$, $F$ (370)

$$\begin{aligned} T_\delta &= \frac{Z+\delta}{\sqrt{V/n}} \sim \mathcal{T}_n(\delta) \\ \sum_1^m (Z_i + \delta_i)^2 &= (Z_1 + \delta)^2 + \sum_2^m Z_i^2 \sim \chi^2_m(\delta) \\ & \frac{\chi^2_m(\delta)/m}{\chi^2_n/n} \sim \mathcal{F}_{m,n}(\delta) \end{aligned}$$

## 5.2 Linear Transformations in Higher Dimensions

### Definitions

**orthogonal** (382) – A 2x2 matrix $A$ is *orthogonal* if its rows are two perpendicular vectors of length 1.
**mean vector** (384) – The *mean vector* of the random vector $\mathsf{X} = (X_1, \ldots, X_n)'$ is the $n \times 1$ vector $\mu = \mu_{\mathsf{X}} = (E(X_1), \ldots, E(X_n))'$.
**covariance matrix** (384) – The *covariance matrix* for the same vector is the $n \times n$ matrix $\Sigma = \Sigma_{\mathsf{X}} = \|\sigma_{ij}\|$.
**non-negative definite** (384) – Every covariance matrix $\Sigma$ is *non-negative definite*, i.e. $\Sigma \ge 0$.
**positive definite** (384) – $\Sigma > 0$.
**orthogonal unit vectors** (386) – $\gamma_1, \ldots, \gamma_n$(with $\gamma_i = (\gamma_{i1}, \ldots, \gamma_{in})'$) such that $\gamma_i'\gamma_i = 1$ for all $1 \le i \le n$, but $\gamma_i'\gamma_j = 1$ for all $i \ne j$.

## Matrix Identities (382)

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad |A| = ad - bc \qquad |A||A^{-1}| = |AA^{-1}| = |I| = 1$$

## Extended Identities (382)

$$\mathsf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} a_{11}X_1 + a_{12}X_2 \\ a_{21}X_1 + a_{22}X_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = A\mathsf{X}$$

$$\text{Jacobian}= \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}^+ = |A^{-1}|^+ = 1/|A|^+ = 1/\begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix}^+$$

## Density of a linearly transformed variable (382)

If $X = A^{-1}Y$, $f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2) \cdot |A^{-1}|^+ = f_{X_1,X_2}(x_1, x_2)/|A^{-1}|^+$

## Mean vectors and covariance matrices of linear transformations (384)

$\mathsf{Y} = A\mathsf{X} + c$ has mean vector $\mu_{\mathsf{Y}} = A\mu_{\mathsf{X}} + c$ and covariance matrix $\Sigma_{\mathsf{Y}} = A\Sigma_{\mathsf{X}}A'$

## Multivariate Normal distribution (385)

For $Z_1, \ldots, Z_n$ independent $N(0, 1)$ rvs,

$$\mathsf{Y} = A\mathsf{Z} + \mu = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{mn} \end{bmatrix}\begin{bmatrix} Z_1' \\ \vdots \\ Z_n' \end{bmatrix} + \begin{bmatrix} \mu_1' \\ \vdots \\ \mu_m' \end{bmatrix} = \begin{bmatrix} \mathsf{a}_i' \\ \vdots \\ \mathsf{a}_m' \end{bmatrix}\begin{bmatrix} Z_1' \\ \vdots \\ Z_n' \end{bmatrix} + \begin{bmatrix} \mu_1' \\ \vdots \\ \mu_m' \end{bmatrix} \sim N(\mu, \Sigma)$$

$\mathsf{Y}$ has the *Multivariate Normal*$(\mu, \Sigma)$ density, determined solely by $\mu_Y$ and $\Sigma_Y = A\Sigma A'$:

$$f_{\mathsf{Y}}(\mathsf{y}) = \frac{1}{|\Sigma|^{1/2}}\frac{1}{(2\pi)^{n/2}}exp(-\tfrac{1}{2}(\mathsf{y}-\mu)'\Sigma^{-1}(\mathsf{y}-\mu))$$

## Orthogonal transformations (386)

$\mathsf{z}=G\mathsf{y}$ where

$$G = \begin{bmatrix} \gamma_1' \\ \vdots \\ \gamma_n' \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} \\ \vdots & & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} \end{bmatrix}$$ is an orthogonal matrix as defined above (note $G'G = I = GG'$).

## Partitioning a multivariate Normal (389)

Let $\mathsf{Y} = A\mathsf{Z} + \mu$ as above, and partition $\mathsf{Y}$, $\mu$, and $\Sigma$ so that

$$\Sigma = AA' = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}[A_1'\ A_2'] = \begin{bmatrix} A_1A_1' & A_2' \\ A_2A_1' & A_2A_2' \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Then $\mathsf{Y} = N(\mu_1, \Sigma_{11})$ and $\mathsf{Y} = N(\mu_2, \Sigma_{22})$.

$\mathsf{Y}_1$ and $\mathsf{Y}_2$ are uncorrelated if and only if $\Sigma_{12} = 0$.

## 5.3 General Transformations in Higher Dimensions

## 5.4 Asymptotics

### Definitions

**convergence in probability** (408) – $U_n \to_p U$ when $P(|U_n - U| \geq \epsilon) \to 0$ as $n \to \infty$, for every $\epsilon > 0$.

**convergence in distribution** (408) – $W_n \to_d W$ when the df $F_n$ of $W_n$ convergest to the df $F$ of $W$ at each point $x$ where $F$ is continuous. We also say that $W_n$ is *asymptotically distributed* as $W$.

### Classical CLT of Probability (408)

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to Z \sim N(0,1)$$

### Finite Sampling CLT (409)

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma\sqrt{1 - \frac{n-1}{N-1}}} \to_d Z \sim N(0,1)$$

### Other applications (409)

**PLT**: $T_n \sim$Binomial$(n, p_n) \to_d T \sim$Poisson$(\lambda)$ when $\lambda_n = np_n \to \lambda$.
**Minimum of Uniforms**: $nU_{n:1} \to_d$Exponential(1).
**Weakest Link Limit Theorem**: $r_n Y_{n:1} \to_d$Weibull$(0, \alpha, \beta)$.
**Discrete Uniform**: For $X_n$ with Discrete Uniform(0,$n$) distribution, $X_n/n \to_d$Uniform(0,1).

### Markov Inequality (409)

For any rv $X$, $P(|X| \geq t) \leq E[|X|]/t$ for all $t > 0$

### Chebyshev Inequality (409)

For $X$ with finite mean $\mu$ and variance $\sigma^2$,
$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$ for all $t > 0$.

### Weak Law of Large Numbers (WLLN) (410)

Let $X_1, X_2, \ldots, X_n$ be iid with finite mean $\mu$. Then
$\bar{X}_n \to_p \mu$.

If, additionally, the $X_i$ have common finite variance $\sigma^2$, the WLLN gives
$S_n^2 \to_p \sigma^2, \quad \hat{\sigma}_n^2 \to_p \sigma^2, \quad$ and $\hat{\sigma}_{on}^2 \to_p \sigma^2$.

### Slutsky's Theorem (411)

For $W_n \to_d W, U_n \to_p a, V_n \to_p b$, and $h$ is continuous at $a$. Then
$U_n W_n + V_n \to_d aW + b$ and $h(U_n) \to_p h(a)$.

The CLT of Statistics thereby follows from the CLT of Probability: $\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \to_d \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to_d N(0,1)$.

### Mann-Wald Theorem (Continuous Mapping Theorem) (412)

For continuous functions $h$, $W_n \to_d W$ implies $h(W_n) \to_d h(W)$.

e.g. $Z_n^2 = [\sqrt{n}(\bar{X}_n - \mu)/\sigma]^2 \to_d N^2(0,1) \sim \chi_1^2$ follows from Mann-Wald with $h(t) = t^2$.

### Delta Method (413)

If $Z_n = \sqrt{n}(\bar{T}_n - \nu) \to_d Z = N(0, \tau^2)$ and $g$ is differentiable at $\nu$, then

$\sqrt{n}[g(\bar{X}_n) - g(\mu)] =_a g'(\nu) \times Z \to_d g'(\nu) \times Z = N(0, [g'(\nu)]^2 \cdot \tau^2)$

### Bivariate CLT of Probability (413)

$$\begin{bmatrix} \bar{Z}_n^X \\ \bar{Z}_n^Y \end{bmatrix} = \begin{bmatrix} \sqrt{n}(\bar{X}_n - \mu_X) \\ \sqrt{n}(\bar{Y}_n - \mu_Y) \end{bmatrix} \to_d \begin{bmatrix} Z_X \\ Z_Y \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \right)$$

### Bivariate Mann-Wald (414)

For continuous functions $h(\cdot,\cdot)$, and $\bar{Z}_n^X$ and $\bar{Z}_n^Y$ as defined above, $h(\bar{Z}_n^X, \bar{Z}_n^Y) \to_d h(Z_X, Z_Y)$ as $n \to \infty$.

e.g. For $Y_i = (X_i - \mu)^2$ having mean $\mu$ and variance$\sigma^2$,

$$\sqrt{n}\begin{bmatrix} \bar{X}_n - \mu_X \\ S_n^2 - \sigma^2 \end{bmatrix} =_a \sqrt{n}\begin{bmatrix} \bar{X}_n - \mu_X \\ \bar{Y}_n - \sigma^2 \end{bmatrix} \to_d N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix} \right)$$

### Two-Dimensional Delta Method (415)

$$\begin{aligned} \sqrt{n}[g(\bar{X}_n, \bar{Y}_n) - g(\mu_X, \mu_Y)] &=_a \left(\nabla g(\mu)\right) \times \begin{bmatrix} Z_X \\ Z_Y \end{bmatrix} = c_1 Z_X + c_2 Z_Y \\ &= \mathsf{c}' \begin{bmatrix} Z_X \\ Z_Y \end{bmatrix} \sim N(0, \mathsf{c}\Sigma\mathsf{c}) = N(0, c_1^2\sigma_X^2 + 2c_1c_2\sigma_{XY} + c_2^2\sigma_Y^2). \end{aligned}$$

### More General CLTs (416)

Weighted average CLT
Liapunov CLT
Lindeberg-Feller CLT
Multivariate CLT

## 5.5 Moment Generating Functions

### Moment Generating Function (mgf) (421)

$M(t) = M_X(t) = E[e^{tX}]$

### Characteristic Function (chf) (421)

$\phi(t) = \phi_X(t) = E[e^{itX}]$

### Cumulant Generating Function (cgf) (421)

$K(t) = K_X(t) = \log \phi(t)$

## Math Facts

$e^{itX} = \cos(tX) + i\sin(tX)$ $|e^{itX}| \leq 1$ for all possible values of $X$ and $t$.

$e^{tX} = 1 + (tX)/1! + \cdots + (tX)^k/k! + \cdots$

## Fundamental Properties of Transforms (422)

Linear changes in $X$
$M_{aX+b}(t) = E[e^{t(aX+b)}] = e^{tb}E[e^{(at)X}] = e^{tb}M_X(at)$
$\phi_{aX+b}(t) = E[e^{it(aX+b)}] = e^{itb}E[e^{i(at)X}] = e^{itb}\phi_X(at)$

MGF of a sum of independent rvs
$M_{X+Y}(t) = M_X(t)M_Y(t)$
$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$
$K_{X+Y}(t) = \log\phi_{X+Y}(t) = \log(\phi_X(t)\phi_Y(t)) = K_X(t) + K_Y(t)$

Uniqueness Theorem
$M_X(\cdot)$ completely determines $P_X(t)$, and vice versa, with qualifications.
$\phi_X(\cdot)$ completely determines $P_X(t)$, and vice versa.

Cramér-Levy Continuiuty Theorem
If $M_{Z_n}(t) \to M_Z(t)$, then $Z_n \to_d Z$, with qualifications.
$\phi_{Z_n}(t) \to \phi_Z(t)$ if and only if $Z_n \to_d Z$ for all t.

## Getting moments from the MGF (423)

$M^{(k)}(0) = E[X^k]$

## Getting moments from the CGF (424)

$K^{(j)}(0) = i^j\kappa_j$
For any rv $X$ with $\mu_4 < \infty$, $\kappa_1 = \mu$ $\kappa_2 = \sigma^2$ $\kappa_3 = \mu_3$ $\kappa_4 = \mu_4 - 3\mu_2^2 = \mu_4 - 3\sigma^4$
For standardized rvs, $\kappa_3 = \gamma_1$ =skewness $\kappa_4 = \gamma_2$ =kurtosis

## Multivariate Moment Generating Functions (429)

$M_{\mathsf{X}}(textsft) = E[e^{textsft'textsfX}] = E[e^{t_1X_1+\cdots+t_nX_n}]$
$\phi_{\mathsf{X}}(textsft) = E[e^{itextsft'textsfX}] = E[e^{it_1X_1+\cdots+it_nX_n}]$
Analogs of the four fundamental properties listed above exist for rvs in $n$ dimensions.

## Multivariate Normal MGF (429)

# Chapter 6

# Classical Statistics

## 6.1 Estimation

### Definitions

**parameter space** (435) – The set $\underline{\Theta}$ of all possible values of the parameter $\theta$ of the model
**statistic** (435) – any function of the data
**unbiased** (435) – If $E_\theta[T(\mathsf{X}) = \tau(\theta)$ for each $\theta$, then $T$ is an unbiased estimator for $\tau(\theta)$
**consistent** (435) – If $T_n(\mathsf{X}) \to_p \tau(\theta)$ for each $\theta$, then $T$ is a consistent estimator for $\tau(\theta)$
**bias** (436) – $E_\theta[T] - \tau(\theta)$
**mean squared error** (MSE) (436) – $\mathrm{Mse}(\theta) = E_\theta[(T - \tau(\theta))^2]$
**method of moments estimators** (437) – Parameter estimators resulting from equating the population moments with the sample moments, and solving for the parameters

### Main Elementary Facts (436)

$Mse(\theta) = Var_\theta(T) + Bias^2(\theta)$
$T_n$ is consistent whenever $\mathrm{Mse}_{T_n}(\theta) \to 0$ as $n \to \infty$,
or whenever $\mathrm{Var}_\theta(T_n) \to 0$ and $\mathrm{Bias}_{T_n}(\theta) \to 0$.

## 6.2 The One-Sample Normal Model

### Definitions

**confidence intervals** (441) – $\bar{X}_n \pm t^* S_n/\sqrt{n}$ provides a $(1-\alpha)$ confidence interval for $\mu$.
**random theoretical confidence interval** (441) – with probability $1-\alpha$, will contain the fixed unknown value $\mu$. This probability is exact for a Normally distributed experiment $X$, and approximately correct for any experiment $X$ having finite mean and variance.
**numerical confidence interval** (441) – We have $1-\alpha$ degree of belief that the non-random numerical value $\bar{x}_n$ did estimate $\mu$ to within an estimated margin for error of $t^* s_n/\sqrt{n}$.
**level-$\alpha$ test** (442) – Reject $H_o : \mu = \mu_o$ in favor of $H_a : \mu \neq \mu_o$ whenever $|T_n(\mu_o)| > t_{n-1}^{(\alpha/2)}$.
**power** (442) – $\mathrm{Power}(\mu, \sigma) = P_{\mu,\sigma}(H_o \text{ is rejected}) = P(|T_{n-1}(\delta)| > t_{n-1}^{(\alpha/2)})$.
**p-value** (444) – chance of the statistic taking on a value at least as extreme as its observed value; numerical evaluation of the strength of evidence against the null hypothesis.
**model equation** (445) – $X_i = \mu + \epsilon_i$, for $1 \le i \le n$, with independent $N(0, \sigma^2)$ rvs $\epsilon_i$.
**location parameter** (445) – $\mu$ in the model equation.
**scale parameter** (445) – $\sigma$ in the model equation.
**errors, residual** (445) – $\epsilon_i = X_i - \hat{\mu} = \text{Observation}_i - \text{Fitted Value}_i$.
**sum of squares** (445) – $SS(\hat{\mu}) = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = SS_{XX}$
**fitted value, least squares estimator** (445) – $\hat{\mu}$ for which $SS_{XX}$ is minimized.
**pth quantile** (446) – $z_p$ such that $P(Z < z_p) = p$.

### Theorem 6.2.1 – One-Sample Normal Theory Building Blocks (441)

For an independent random sample $X_1, \ldots, X_n$ from the $N(\mu, \sigma^2)$ distribution,

$Z_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma \sim N(0,1)$

$\frac{S_n^2}{\sigma^2} = \frac{1}{\sigma^2} \cdot \frac{SS_{XX}}{n-1} \sim \frac{\chi^2_{n-1}}{n-1}$

$\bar{X}_n$ and $S_n^2$ are independent rvs.

$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{SS_{XX}/(n-1)}} \sim \mathcal{T}_{n-1}$.

### Kantorovich Inequality (449)

If $P(a \le X \le b) = 1$ for some $a \ge 0$, then $1 \le E[X] \cdot E[1/X] \le (a+b)^2/(4ab)$.

### Jensen Inequality (449)

$g(\mu) \le E[g(X)]$, for convex $g$ on the interval $I$, where $g$ has mean $\mu$ in the interior of the interval $I$.

e.g. For $g(X) = X^2$, we obtain $E[X^2] - \mu^2 = \sigma^2 \ge 0$.

## 6.3 The Two-Sample Normal Model

### Definitions

**variance ratio** (457) – $\Delta^2 = \sigma_2^2/\sigma_1^2$.

**paired differences** (458) – For an independent sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from a basic $(X, Y)$-experiment with joint Bivariate Normal distribution, $D_i = X_i - Y_i$ is the *paired difference.*

**efficiency of estimators** (461) – $\mathcal{E}(\bar{\mu}_n, \hat{\mu}_n) = \frac{Var[\hat{\mu}_n]}{Var[\bar{\mu}_n]} \to \mathcal{E}(\bar{\mu}, \hat{\mu})$.

### Equal Variances (456)

For independent samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ from $N(\mu, \sigma^2)$ and $N(\nu, \sigma^2)$, and $\theta = \mu - \nu$,

$\hat{\theta} = \bar{X}_m - \bar{Y}_n \sim N(\theta, \frac{m+n}{mn}\sigma^2)$.

$S^2 = \frac{m-1}{m+n-2}S_X^2 + \frac{n-1}{m+n-2}S_Y^2 = \frac{1}{m+n-2}(SS_{XX}^2 + SS_{YY}^2) \sim \frac{\sigma^2}{m+n-2}\chi^2_{m+n-2}$.

It follows from the independence of $\hat{\theta}$ and $S^2$ that
$T = \sqrt{\frac{mn}{m+n}} \frac{(\bar{X}_m - \bar{Y}_n) - (\mu - \nu)}{S_{m,n}} = \sqrt{\frac{mn}{m+n}} \frac{\hat{\theta} - \theta}{S_{m,n}} \sim$Student-$\mathcal{T}_{m+n-2}$.

### Unequal variances (456)

For independent samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ from $N(\mu, \sigma_1^2)$ and $N(\nu, \sigma_2^2)$, and $\theta = \mu - \nu$,

$\hat{\theta} = \bar{X}_m - \bar{Y}_n \sim N(\theta, \frac{m+n}{mn}\bar{\sigma}^2)$, where $\bar{\sigma}^2 = \frac{n\sigma_1^2 + m\sigma_2^2}{m+n}$.

$\bar{S}^2 = \frac{nS_X^2 + mS_Y^2}{m+n}$.

By the CLT (provided only that the basic $X$ and $Y$ have finite variances,
$\bar{T} = \sqrt{\frac{mn}{m+n}} \frac{(\bar{X}_m - \bar{Y}_n) - (\mu - \nu)}{\bar{S}_{m,n}} = \frac{(\bar{X}_m - \bar{Y}_n) - (\mu - \nu)}{\sqrt{\frac{1}{m}S_X^2 + \frac{1}{n}S_Y^2}} \to_d N(0,1)$.

### Known Variance Ratio ($\Delta^2$) (457)

The pooled estimator above now satisfies $S^2 \sim \frac{\sigma_1^2}{m+n-2}(\chi^2_{m-1} + \Delta^2 \chi^2_{n-1})$, but a more helpful statistic is $S^2(\Delta) = \frac{m-1}{m+n-2} S_X^2 + \frac{n-1}{m+n-2} \frac{1}{\Delta^2} S_Y^2 \sim \frac{\sigma_1^2}{m+n-2}(\chi^2_{m+n-2})$.

The corresponding $T$-statistic is
$T(\Delta_o) = \sqrt{\frac{mn}{n+m\Delta_o^2}}[(\bar{X}_m - \bar{Y}_n) - (\mu - \nu)]/S_{m,n}(\Delta_o) \sim \mathcal{T}_{m+n-2}$.

### Matched Pairs $T$-test (458)

As defined above, $D = X - Y \sim N(\mu_D, \sigma_D^2) = N(\theta, \sigma_D^2)$, where the variance is given by
$\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$.

From Theorem 6.1.2, we have
$Z_D = \sqrt{n}(\bar{D}_n - \theta)/\sigma_D \sim N(0,1)$,
$S_D^2 = \frac{1}{n-1}\sum_{i=1}^2 (D_i - \bar{D})^2 \sim \frac{\sigma_D^2}{n-1}\chi^2_{n-1}$,

It follows from the independence of $\bar{D}$ and $S_D^2$ that
$T_D = \sqrt{n}(\bar{D}_n - \theta)/S_{D,n} \sim \mathcal{T}_{n-1}$.

### Normal Theory Building Blocks (461)

The optimal choice for weighting two samples is that which minimizes the variance of $\hat{\mu}$.
The natural estimator of the variance employs the same weighting used for $\hat{\mu}$.

## 6.4 Other Models

### Examples

Bernoulli (466)
Poisson (467)
NegBiT (469)

# Chapter 7

# Estimation, Principles and Approaches

## 7.1 Sufficiency, and UMVUE Estimators

### Definitions

**sufficient statistic** (473) – $T = T(\mathsf{X})$ is a sufficient statistic for $\theta$ (or for the family $\mathcal{P} = \{P_\theta : \theta \in \underline{\Theta}\}$) when the conditional distribution $P_{\mathsf{X}|T=t}(\cdot)$ does not depend on $\theta$.
**ancillary statistic** (474, 486) – A statistic $V = V(\mathsf{X})$ is called ancillary if the distribution of $V$ does not depend on $\theta$.
**zero function** (476) – $h(t) = 0$ for all $t$.
**complete family of distributions** (476, 484) – An arbitrary family $\mathcal{P}_t$ of the possible distributions (of some rv $T$) is called complete if the condition $E_P[h(T)] = 0$ for every possible distribution of $P$ of $T$ in the family $\mathcal{P}_t$ implies that $h$ can only be the zero function.
**uniform minimum unbiased estimator (UMVUE)** (476) – Suppose $T$ is an unbiased estimator of $\tau(\theta)$ for which $\mathrm{Var}_\theta[T] \le \mathrm{Var}_\theta[U]$ for all $\theta$, for any other unbiased estimator $U$ of $\tau(\theta)$. Then $T$ is a UMVUE of $\tau(\theta)$.
**minimally sufficient statistic** (478) – In general, the statistic $T(X)$ is said to be minimally sufficient in some model if any other sufficient statistic $\tilde{T} = \tilde{T}(\mathsf{X})$ for this model necessarily satisfies a relationship of the form $T = \tilde{h}(\tilde{T})$ for some function $\tilde{h}$.
**common support** (480) – The *support* of $P_\theta$ is $\{\mathsf{x} : p_\theta(\mathsf{x}) > 0\}$. A family of densities for which the support of each density is the same is said to have *common support*.

### Sufficiency Principle (474)

If $T = T(\mathsf{X})$ is a sufficient statistic for $\theta$ in the model, then to estimate any parameter of the form $\tau(\theta)$ one should only use estimators that are functions of $T$.

### Ancillarity Principle (474)

Basing an analysis on the conditional distribution given the ancillaries might well access useful information about the underlying distribution. (Often useful, it is less than compelling; ancillaries are not unique.)

### Fisher-Neyman Factorization Theorem (476)

$T(X)$ is sufficient for $\theta$ if and only if the model distributions can be factored as
$p_\theta(\mathsf{x}) = g_\theta(t) \times h_t(\mathsf{x}) = g_\theta(t) \times h(\mathsf{x})$
$f_\theta(\mathsf{x}) = g_\theta(t) \times h_t(\mathsf{x}) = g_\theta(t) \times h(\mathsf{x})$

### Rao-Blackwell (476)

Let $U = U(\mathsf{X})$ be unbiased for $\tau(\theta)$, and suppose $T(X)$ is sufficient for $\theta$. Then
$V = V(T) = E(U|T)$
a) is a function of the sufficient statistic,
b) is unbiased for $\tau(\theta)$ since
$E_\theta(V) = E_\theta(m(T)) = E_\theta(E(U|T)) = E_\theta(U) = \tau(\theta)$, and
c) is at least as good as $U$ since
$Var_\theta(V) = Var_\theta(m(T)) \le Var_\theta(E(U|T)) + E_\theta(Var(U|T)) = Var_\theta(U)$.

### Lehmann-Scheffé (unique UMVUEs) (476)

Let $T$ be sufficient for $\theta$ and complete, and $V = V(T)$ be any unbiased estimator of $\tau(\theta)$ with finite variance. Then

$V$ is the *unique UMVUE* of $\tau(\theta)$. (Often a computation of the form $V = E(U|T)$ will be useful.)

### Likelihood Ratios and Sufficiency (481)

Within the common support model, $T(\mathsf{X})$ is sufficient for $\theta$ if and only if each likelihood ratio is a function of this $T$.

### Examples

Sufficiency geometrically (normal spheres)(477)
Condition on Ancillary Statistics (481)

## 7.2 Completeness, and Ancillary Statistics

### Math Fact 1 (infinite series) (484)

$h(x) = \sum_{k=0}^{\infty} a_k x^k = 0$ for all $x$ in some interval iff $a_k = 0$ for all $k$.

### Math Fact 2 (only the zero function) (484)

$\int_{(-\infty,x]} h(t)\,dt = 0$ for all $x$ iff $h(x) = 0$ for all $x$.

### Basu (487)

If $T = T(\mathsf{X})$ is sufficient and complete for $\theta$, then every ancillary statistic $V(\mathsf{X})$ is independent of $T$.

### Sufficiency and Completeness Relative to Inclusion (487)

If $T = T(\mathsf{X})$ is sufficient for $\mathcal{P}$, then $T$ is sufficient for the smaller $\mathcal{P}_o$.
if the family $\mathcal{P}_o^T$ is complete, then the bigger family $\mathcal{P}^T$ is complete.

### Examples

Uniform$(0,\theta)$, Uniform $(\theta,\theta)$ (485) – complete and sufficient statistics
Uniform$(0,\theta)$ (486) – ancillary statistics
Exponential$(\theta)$ (486) – ancillary statistics
Poisson UMVUEs (487)

## 7.3 Exponential Families, and Completeness

### Definitions

**exponential family** (492) – A family $\mathcal{P}_o = \{P_\theta : \theta \in \underline{\Theta}_o\}$ of distributions that can be written in the format $c(\theta)e^{\sum_{i=1}^{\kappa} a_i(\theta)T_i(x)}h(x)1_D(x)$ where $c(\theta)$ depends only on $\theta$, $h(x) > 0$ iff $x \in D$. The family has common support $D$.

**Order of** $\mathcal{P}_o$ (492) – the minimum $\kappa$ for which such an expression of an exponential family is possible.

**affinely independent over** $D$ (492) – means that $\sum_1^{\kappa} c_i T_i(x) = c_o$ for a.e. $x \in D$ implies that all $c_i = 0$

**minimal over** $D$ (492) – $\mathsf{T}$ that satisfies affine independence

**affinely independent over** $\underline{\Theta}_o$ (492) – means that $\sum_1^{\kappa} c_i a_i(\theta) = c_o$ for $\theta \in \underline{\Theta}_o$ implies that all $c_i = 0$

**minimal over** $\underline{\Theta}_o$ (492) – $\mathsf{a}$ that satisfies affine independence

**natural parameter space** ($\eta$) (492) – For $\mathcal{P}_o$ with order $\kappa$ and $\mathsf{T}$ and $\mathsf{a}$ both minimal, replace $\mathsf{a}(\theta)$ with $\eta = (\eta_1, \ldots, \eta_\kappa)'$. Rewrite the natural parameter space as $e^{-d(\eta)+\sum_{i=1}^{\kappa} \eta_i T_i(x)}h(x)1_D(x)$, where $d(\eta) = log(\int_D e^{\sum_{i=1}^{\kappa} \eta_i T_i(x)}h(x)\,dx)$. $\mathcal{N} = \{\eta : d(\eta) > \infty\}$ is the *natural parameter space.*

**full exponential family** (492) – $\mathcal{P}_{\mathcal{N}} = \{P_\eta : \eta \in \mathcal{N}\}$ for $\mathcal{N}$ as just defined.

**curved exponential family** (495) – situations where the parameters in the natural parameter space are subject to specified relationships (e.g. Normal$(\theta, \theta)$ or Normal $(\theta, \theta^2)$).

### Natural Parameter Space Properties (492)

1) $\mathcal{N}$ is a convex set and $d(\cdot)$ is convext on $\mathcal{N}$.
If $E_\eta[g(\mathsf{X}] < \infty$ for all $\eta \in \mathcal{N}$, then
2) all derivatives of $\phi(\cdot)$ exist and
3) $\phi(\eta) = E_\eta[g(\mathsf{X})]$ can be differentiated "under the integral sign" on the interior of $\mathcal{N}$.

### Theorem 7.3.1: Sufficient and complete statistic for exponential families (493)

Let $X_1, ..., X_n$ be iid, with each distribution in a family of order $\kappa$ given by

$$c(\theta)e^{\sum_{i=1}^{\kappa} a_i(\theta)T_i(x)}h(x)1_D(x)$$

**(A)** For $1 \le i \le \kappa$, define $T_i(\mathsf{X}) = \sum_{j=1}^{n} T_i(X_j)$, and then define the statistic $\mathsf{T} = \mathsf{T}(\mathsf{X})$.
This $\mathsf{T}$ is sufficient for $\mathcal{P}_{\mathcal{N}}$.

**(B)** If $\mathcal{N}_o = \{(a_1(\theta), ..., a_n(\theta)) : \theta \in \Theta_o\}$ contains a $\kappa$-dimensional subset $\mathcal{H}$, then
$\mathsf{T}$ is both minimally sufficient and complete for $\mathcal{P}_{\mathcal{N}_l} = \mathcal{P}$.

**(C)** The sufficient statistic $\mathsf{T}$ necessarily satisfies
$\mathsf{T}$ is distributed as an exponential family with the same $a_i(\theta)$'s as $X$.

**(D)** Suppose $\kappa$ is fixed. The conditional distribution of $(T_1|T_2 = t)$ is again an exponential family.

**(E)** The conditional distribution $P_{\mathsf{X}|\mathsf{T}=\mathsf{t}}$ takes the form

$$\frac{c^n(\theta)e^{a(\theta)t} \times \prod_{j=1}^{n} h(x_j)}{c^n(\theta)c^{a(\theta)t} \times h^*(t)} = \frac{\prod_{j=1}^{n} h(x_j)}{h^*(t)}$$

### Theorem 7.3.2: Sufficiency and completeness of one-to-one transformations (494)

The outcome of some $X$-experiment is modeled as being distributed according to one of the distributions $P_\theta$ in some family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. Suppose $T = T(X)$ is sufficient and complete for $\theta$. Let $U = U(T)$ denote any one-to-one function of $T$. The $U$ is sufficient and complete for $\theta$. (Also, $T, \theta$, and $U$ may all denote vectors.)

## 7.4 The Cramér-Rao Bound

### Definitions

**likelihood of** $\theta$ (503) – $L(\theta;\mathsf{x}) = p_\theta(\mathsf{x})$ OR $f_\theta(\mathsf{x})$. In the case of independent rvs $X_k$, $L(\theta;\mathsf{x}) = \prod_{k=1}^{n} L(\theta;x_k)$.

**maximum likelihood estimator** (503) – the value of $\theta$ in the parameter set $\underline{\Theta}$ that maximizes the likelihood $L(\theta;\mathsf{x})$.

**log-likelihood** (503) – $\ell(\theta;\mathsf{x}) = \log L(\theta;\mathsf{x}) \overset{iid}{=} \sum +k = 1^n \ell(\theta;x_k)$

**score function** (503) – $\dot{\ell}(\theta;\mathsf{x}) = \frac{\partial}{\partial\theta}\log p_\theta(\mathsf{x})$ OR $\frac{\partial}{\partial\theta}\log f_\theta(\mathsf{x})$

**likelihood equation** (503) – $\dot{\ell}(\theta;\mathsf{x}) = 0$; the MLE is commonly found by solving this equation.

**fisher information** (504) – $\mathcal{I}_n(\theta) = E_\theta\{[\dot{\ell}_n(\theta;\mathsf{X})]^2\}$.

**alt fisher information** (504) – $\bar{\mathcal{I}}_n(\theta) = E_\theta\{-\ddot{\ell}_n(\theta;\mathsf{X})\}$. this is equal to the fisher information for "regular" distributions

**information per observation** (504) – the Fisher information as calculated from one obeservation $x_k$ of the data

**theoretical score function** (504) – $\dot{\ell}_n(\theta;\mathsf{X})$

**numerical score function** (504) – $\dot{\ell}_n(\theta;\mathsf{x})$

**relative efficiency** (506) – $\mathcal{E}_\theta(\tilde{T},T) = \dfrac{Mse_\theta[T]}{Mse_\theta[\tilde{T}]}$

**absolute efficiency** (506) – $\mathcal{E}_\theta(\tilde{T},T) = CRB_n(\theta)/Mse_\theta[\tilde{T}]$

**asymptotic relative efficiency** (506) – the limit of absolute efficiencey as $n \to \infty$

### Elementary Fact (505)

$\mathrm{Cov}(U,V) = E[(U-0)(V-\mu_V)]$ when $E[U]=0$.

### Cramér-Rao, Information Inequality, Equality Corollary (504)

**(A)** Let the model distributions $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ have common support. If the equation $1 = E_\theta(1)$ can be differentiated under the integral sign, then the score function has mean 0; that is
$E_\theta[\dot{\ell}(\theta,X)] = 0$.

**(B)** If the equation $1 = E_\theta(1)$ can be differentiated under the integral sign twice, then the score function has variance $\bar{\mathcal{I}}_n(\theta)$. That is (using (A)),
$\mathcal{I}_n(\theta) = E_\theta[(\dot{\ell}_n(\theta,X))^2] = Var_\theta[\dot{\ell}_n(\theta,X)] = \bar{\mathcal{I}}_n(\theta) = E_\theta[-\ddot{\ell}_n(\theta,X)]$

**(C) (Information Inequality)** Mainly, let $T = T(X)$ denote any statistic of interest for which $E_\theta[|T(X)|] < \infty$, and define $\tau(\theta) = E_\theta[T(X)]$. Suppose
$\tau(\theta)$ has a derivative $\tau'(\theta)$ on the interior of $\Theta$, and both
$\tau(\theta)$ and $1 = E_\theta(1)$ can be differentiated once under the integral.

Then a bound on the variance of any such unbiased estimator $T(X)$ of $\tau(\theta)$ is
$Var_\theta[T(X)] \geq \frac{[\tau'(\theta)]^2}{I_n(\theta)}$
Thus, we have a standard which no unbiased estimator can surpass.

**(D) (Equality Corollary)** Equality is achieved in (C) for a particular estimator $T(X)$ of $\tau(\theta)$ only if $\dot{\ell}(\theta,X) = A_n(\theta)[T(X) - \tau_o(\theta)]$ for some constant $A_n(\theta)$,
and in this case the CR-bound of (C) becomes
$Var_\theta[T(X)] = \left|\frac{[\tau_o'(\theta)]}{A_n(\theta)}\right|$.

Moreover, when this relationship holds,
$\mathcal{I}_n(\theta) = |\tau_o'(\theta)A_n(\theta)|$
This format can hold for at most one function $\tau_o(\cdot)$(and linear variations on it).

### Theorem 7.4.2: UMVU Estimators of $\tau(\theta)$ (506)

$X$ is distributed according to one of the distributions $P_\theta$ in $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$.

(a) Let $V$ be a UMVUE of $\tau(\theta)$ that has finite variance for all $\theta$.
Then $V$ is unique.

(b) Let $V$ be unbiased for $\tau(\theta)$ that has finite variance within the model. Let $U$ denote any statistic having finite variance within the model.
Then $V$ is UMVUE of $\tau(\theta)$ iff $Cov_\theta[U, V] = 0$ for all $U$ with $E_\theta[U] = 0$ for all $\theta$.

### Example

Poisson($\nu \cdot a_i$) regression (508)
Truncation family (509)
Power family (509)
Non-regularity of Uniform$(0, \theta)$ (510)

## 7.5 Maximum Likelihood Estimators

### Definitions

**maximum likelihood estimator** (513) – $\hat{\theta}_n(\mathsf{x})$ is the value of $\theta$ in the parameter set $\underline{\Theta}$ that maximizes the likelihood $L(\theta; \mathsf{x})$.
**theoretical maximum likelihood estimator** (513) – $\hat{\theta}_n(\mathsf{x})$, the MLE when evaluated at the random unperformed experimental outcome $\mathsf{X}$.
**sufficiency** (513) – If $T(X)$ is sufficient for $\theta$, then any MLE $\hat{\theta}_n$ is a function of $T$.
**invariance** (513) – If $\hat{\theta}_n$ is the MLE of $\theta$, then the MLE of $\tau(\theta)$ is $\tau(\hat{\theta}_n)$.
**general asymptotic normality and $a_n$-consistency** (529) – $a_n(\tilde{\theta}_n - b_n(\theta)) \rightarrow_d N(-, v_\theta^2)$
**asymptotic normality and $\sqrt{n}$-consistency** (513) – If an iid model has nice smoothness properties, then $\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_d N(0, \mathcal{I}^{-1}(\theta))$ for the Fisher information in one $X$.
**fisher information for a vector** (516) – $\bar{\mathcal{I}}_n(\theta) = \begin{bmatrix} E_\theta[-\frac{\partial^2}{\partial\alpha^2}\ell_n(\theta; \mathsf{X})] & E_\theta[-\frac{\partial^2}{\partial\beta\,\partial\alpha}\ell_n(\theta; \mathsf{X})] \\ E_\theta[-\frac{\partial^2}{\partial\alpha\,\partial\beta}\ell_n(\theta; \mathsf{X})] & E_\theta[-\frac{\partial^2}{\partial\beta^2}\ell_n(\theta; \mathsf{X})] \end{bmatrix}$
**(alternative for)** (516) – $\mathcal{I}_n(\theta) = \begin{bmatrix} E_\theta[(\frac{\partial}{\partial\alpha}\ell_n(\theta; \mathsf{X}))^2] & E_\theta[(\frac{\partial}{\partial\beta}\ell_n(\theta; \mathsf{X}))(\frac{\partial}{\partial\alpha}\ell_n(\theta; \mathsf{X}))] \\ E_\theta[(\frac{\partial}{\partial\alpha}\ell_n(\theta; \mathsf{X}))(\frac{\partial}{\partial\beta}\ell_n(\theta; \mathsf{X}))] & E_\theta[(\frac{\partial}{\partial\beta}\ell_n(\theta; \mathsf{X}))^2] \end{bmatrix}$
**regularity conditions** (517) – The previous alternative forms the fisher information matrix are equivalent under certain regularity conditions (see Theorem 7.5.1). Under these conditions, the theoretical score function has mean 0 and variance $\mathcal{I}(\theta)$.
**expected information** (517) – $\mathcal{I}(\theta)$
**information random variable** (517) – $I_n(\theta) = -\dfrac{1}{n}\ddot{\ell}_n(\theta; \mathsf{X}) = \bar{V} \rightarrow_p \mu_V = \bar{\mathcal{I}}$
**observed information** (517) – $\hat{I}_n = I_n(\hat{\theta}_n) = -\frac{1}{n}\sum_{k=1}^n \ddot{\ell}(\hat{\theta}_n; X_k)$
**estimated information** (518) – $\mathcal{I}(\hat{\theta}_n)$
**Newton-Raphson iteration method** (522)

### Theorem 7.5.1: Asymptotic properties of MLEs (518)

**Regularity Conditions** **(519)**
Assumptions that suffice (even for an $r$-dimensional $\theta$) are:
(i) $Z_n(\theta_o) \rightarrow_d N(0, \mathcal{I}(\theta_o))$ and $I_n(\theta_o) \rightarrow_p \bar{\mathcal{I}}(\theta_o) = \mathcal{I}(\theta_o) > 0$ is finite.
(ii) $\hat{\theta}_n$ satisfies the likelihood equation.
(iii) $\hat{\theta_n} \rightarrow_p \theta_o$.
(iv) The mean value theorem is applicable.
(v) $|D_n| = |I_N(\hat{\theta}_n^*) - I_n(\theta_o)| \leq \sup[|I_N(\theta') - I_n(\theta_o)| : |\theta' - \theta_o| \leq |\hat{\theta}_n - \theta_o|] \rightarrow_p \mathsf{0}$.
(vi) distributions $P_\theta$ have common support.

**Conclusions** **(518)**

Consider an iid model that satisfies these regularity conditions. Then $\hat{\theta}_n$ and $\hat{I}_n = I_n(\hat{\theta}_n)$ satisfy:

(i) $\sqrt{n}(\hat{\theta}_n - \theta) =_a \mathcal{I}^{-1}(\theta_o) Z_n(\theta_o) = \mathcal{I}^{-1}(\theta_o)[\frac{1}{\sqrt{n}} \sum_{k=1}^n \dot{\ell}(\theta_o; X_k)$

(ii) $\quad \to_d \mathcal{I}^{-1}(\theta_o) Z(\theta_o) \sim \mathcal{I}^{-1}(\theta_o) N(0, \mathcal{I}(\theta_o)) \sim N(0, \mathcal{I}^{-1}(\theta_o))$

(iii) $\hat{I_n} = I_n(\hat{\theta}_n) \to_p \mathcal{I}(\theta_o) > 0$ for the observed information $\hat{I}_n = I_n(\hat{\theta}_n)$.

(iv) $\sqrt{n}(\hat{\theta}_n - \theta) \cdot \hat{I}_n^{1/2} \to_d N(0, I_m)$ (where $I_m$ is the $m \times m$ identity matrix).

(v) $\hat{\theta}_n \pm \frac{1}{\sqrt{n}} z^{(\alpha/2)} \hat{I}_n^{-1/2}$ (for an $m = 1$ dimensional parameter $\theta$).

## Theorem 7.5.2: Asymptotic properties of MLE vectors (519)

**Regularity Conditions** **(519)**

Assuming the following regularity conditions:

**(N0)**

(i) Observations $X_1, ..., X_n$ are iid as one of the distinct distributions $P_{\theta_o}$ in $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ defined by densities $f_\theta(\cdot)$ or mass functions $p_\theta(\cdot)$.

(ii) The parameter vector $\theta$ is $m$-dimensional with $\underline{\Theta}$ a subset of $R_m$.

(iii) The support ($\{x : f_\theta(x) > 0\}$ or $\{x : p_\theta(x) > 0\}$) does not depend on $\theta$.

**(N1)**

(i) All partials have mean 0 under $\theta_o$; thus, $E_{\theta_o}[\dot{\ell}_i(\theta_o; X)] = 0$ for all $i$.

(ii) All partials have finite variances under $\theta_o$; thus $E_{\theta_o}\{[\dot{\ell}_i(\theta_o; X)]^2\} < \infty$ for all $i$.

(iii) $\bar{I}(\theta_o) = I(\theta_o) > 0$.

**Also:** $\Theta$ contains an open neighborhood $\mathcal{H}$ of the true parameter value $\theta_o$ in which:

**(N2)** For all $x$, all second order partial derivatives of $\ell(\theta; x)$ are continuous in $\theta$ and the magnitude of each is bounded by a function of $M(x)$ for which $E_{\theta_o}[M(x)] < \infty$.

**(N3)** For all $x$, all third order partial derivatives of $\ell(\theta; x)$ are continuous in $\theta$ and the magnitude of each is bounded by a function of $M(x)$ for which $E_{\theta_o}[M(x)] < \infty$.

**Conclusions** **(519)**

Let $\hat{\theta}_n$ be a solution of the likelihood equation for which $\hat{\theta}_n \to_p \theta_o$, as required.

Suppose that either all of the regularity conditions hold, or that the assumptions for the vector version hold. Then the conclusions of Theorem 7.5.1 still hold.

## Newton-Raphson Iteration Method (522)

Newton=Raphson provides iterated solutions to the liklihood equations. The method works extremely well if $H'(\theta) > 0$ (or $H'(\theta) < 0$) over the entire region in question. The solution is found through iteration of a one-term Taylor series approximation.

# 7.6 Procedures for Location and Scale Models

## Definitions

**location and scale model equation** (545) – $Y_k = \theta + \tau \cdot W_k$ for $1 \le k \le n$, where $W_1, \ldots, W_n$ are independent with a fixed and knownd df $F_o(\cdot)$.

**location invariance** (545) – $C(\mathsf{Y}) = C(\theta + \tau \mathsf{W}) = \theta + \tau C(\mathsf{W})$

**scale invariance** (545) – $V(\mathsf{Y}) = V(\theta + \tau \mathsf{W}) = \tau V(\mathsf{W})$

**pivot** (545) – $T = T(\mathsf{Y}) = \sqrt{n}\dfrac{C(\mathsf{Y}) - \theta}{V(\mathsf{Y})} = \cdots = \sqrt{n}\dfrac{C(\mathsf{W})}{V(\mathsf{W})} = T(\mathsf{W})$ is a *pivot* for $\theta$ on which confidence intervals can be based. Likewise, $U = U(\mathsf{Y}) = V/\tau = V(\mathsf{W})$ is a *pivot* for $\tau$.

### Natural Estimators of Location (545)

"The examples of location estimators listed below are seen to behave in a very pleasing fasion in the context of a location-scale model. These estimators of location (or center include)"

$\bar{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_k = \frac{1}{n}\sum_{k=1}^{n}(\theta + \tau W_k) = \theta + \tau \cdot \bar{W}_n$

$\ddot{Y}_n = \text{median}_{1\le k\le n} Y_k = \text{median}_{1\le k\le n}(\theta + \tau W_k) = \theta + \tau \cdot \ddot{W}_n$

$\frac{1}{2}(Y_{n:1} + Y_{n:n}) = \theta + \tau \cdot \frac{1}{2}(W_{n:1} + W_{n:n})$

### Natural Estimators of Scale (545)

"The examples of scale estimators listed below are seen to behave in a very pleasing fashion in the context of a location-scale model. These estimators of scale (or dispersion, or variation) include"

$\bar{D}_n = \frac{1}{n}\sum_{k=1}^{n} |Y_k - \bar{Y}_n| = \frac{1}{n}\sum_{k=1}^{n} |(\theta + \tau W_k) - (\theta + \tau \bar{W}_n)|$

$= \tau \cdot \frac{1}{n}\sum_{k=1}^{n} |W_k - \bar{W}_n|$

$S_n = [\frac{1}{n-1}\sum_{k=1}^{n} |Y_k - \bar{Y}_n|^2]^{1/2} = [\frac{1}{n-1}\sum_{k=1}^{n} |(\theta + \tau W)k) - (\theta + \tau \bar{W}_n)|^2]^{1/2}$

$= \tau \cdot [\frac{1}{n-1}\sum_{k=1}^{n} |W_k - \bar{W}_n|^2]^{1/2}$

$Y_{n:n} - Y_{n:1} = (\theta + \tau W_{n:n}) - (\theta + \tau W_{n:1}) = \tau \cdot (W_{n:n} - W_{n:1})$

## 7.7 Bootstrap Methods

## 7.8 Bayesian Methods

### Distributions

**prior distribution** (571) – $\pi(\theta)$ – distribution of model parameter by current knowledge.

**posterior distribution** (571) – $\pi(\theta|\mathsf{x})$ – distribution upon which all Bayesian statistical procedures will be based: $f_{\theta|\mathsf{X}=\mathsf{x}}(\theta) = \dfrac{f(\mathsf{x}|\theta)\pi(\theta)}{f_{\mathsf{X}}(\mathsf{x})} = \dfrac{f(\mathsf{x}|\theta)\pi(\theta)}{\int_{\underline{\Theta}} f(\mathsf{x}|\theta')\pi(\theta')d\theta'}$

**predictive/marginal distribution** (571) – $f_{\mathsf{X}}(\mathsf{x}) = \int_{\underline{\Theta}} f(\mathsf{x}|\theta')\pi(\theta)d\theta'$

### Moments

**posterior mean** (571) – $E(y|x) = m(x) =$ mean of posterior distribution $\pi(\theta|\mathsf{x})$; often can be expressed as a weighted average of the prior mean and the data mean.

**posterior mean** of $\tau(\theta)$ (581) – $\hat{\tau}_B = \hat{\tau}_\pi(\mathsf{x}) = E_\pi(\tau(\theta)|\mathsf{x})$; minimizes $E_\pi(L(\hat{\tau}_B, \tau)|\mathsf{x}) =$ posterior risk with respect to $\pi(\cdot)$ (generally the loss function used is the squared error loss, making this equivalent to minimizing the posterior variance).

**posterior variance** (571) – $\text{Var}(y|x) = v^2(x) =$ variance of posterior distribution $\pi(\theta|\mathsf{x})$.

### Risk

**posterior risk** for $\pi$ (572) – $\text{Var}_\pi(\tau(\theta)|\mathsf{x})$, evaluated at $\hat{\tau}(\mathsf{x}) = E_\pi[\tau(\theta)|\mathsf{x}]$ that minimizes $E_\pi[(\tau(\theta) - \hat{\tau}(\theta))^2|\mathsf{x}]$ among all $\hat{\tau}(\mathsf{x})$.

**risk function** (581) – $R(\hat{\tau}, \theta) = E_\theta[L(\hat{\tau}, \tau(\theta))]$

**$\pi$-averaged risk** of $\hat{\tau}$ (581) – $R_\pi(\hat{\tau}) = E_\pi[R(\hat{\tau}, \theta)] = \int R(\hat{\tau}, \theta)\pi(\theta)d\theta$

## Extra Definitions

$1-\alpha$ **credible set** (574) – $B_{\mathsf{x}}$ s.t. the probability $\tau(\theta)$ is in $B_{\mathsf{x}} = 1-\alpha$, under the posterior distribution. When $B_{\mathsf{x}}$ is an interval, call this a $1-\alpha$ credible interval for $\tau(\theta)$.
$1-\alpha$ **credible interval** for $\theta$ (574) – $I_{\mathsf{x}} = \tau^{-1}(B_{\mathsf{x}})$
**highest probability density intervals** (HPD intervals) (574) – shortest intervals $I_{\mathsf{x}}$ are where the density takes its highest values.
**Bayes estimator** of $\tau(\theta)$ (571) – see posterior mean of $\tau(\theta)$.
**Bayesian interval** for $\theta$ (571) – see $1-\alpha$ credible interval.

## Bayesian Hypothesis Testing (572)

The level $\alpha$ **Bayesian** test of $\underline{\Theta_o}$ vs. $\underline{\Theta_o}^c$ is based on the following hypothesis testing decision rule:
Reject $\underline{\Theta_o}$ if $P_\pi(\underline{\Theta_o}|\mathsf{X}{=}\mathsf{x}) \leq \alpha$

Under this rule, we can define a **Bayesian** p-value as the probability of rejecting the null hypothesis under the (null hypothesis of the) posterior distribution:
$P_\pi(\underline{\Theta_o}|\mathsf{X}{=}\mathsf{x})$

Likewise, we can define **Bayesian** power as the probability of rejecting the null hypothesis under the (alternative hypothesis of the) posterior distribution:
$P_{\tilde{\pi}}(\Theta_o|\mathsf{X}{=}\mathsf{x})$

## Bayesian Sufficiency (572)

If $\pi(\theta|\mathsf{x})$ depends on $\mathsf{x}$ only through $t = T(\mathsf{X})$, then $T$ is *Bayesian sufficient* for $\theta$ (and we can replace $\mathsf{X}$ with $T$ in $\pi(\theta|\mathsf{x})$ above).
If $T$ is sufficient for $\mathcal{P}$, then $T$ is Bayesian sufficient for $\theta$.

## Conjugate priors

**Binomial-Beta** (573) - $Y_i$ ~Binomial$(n,\theta)$ and $\pi$ ~Beta$(\alpha,\beta)$ have posterior distribution Beta$(\alpha^*,\beta^*)$, where $\alpha^* = y+\alpha, \beta^* = n+\beta-y$.
**Gamma-Gamma** (576) - $Y_i$ ~Gamma$(r,\nu)$ and $\pi$ ~Gamma$(r_o,c)$ have posterior Gamma$(r_o^*,c^*)$, where $r_o^* = nr + r_o, c^* = t + c = \sum y_i + c$.
**Poisson-Gamma** (586) - $Y_i$ ~Poisson$(\lambda)$ and $\pi$ ~Gamma$(r,\nu)$ have posterior distribution Gamma$(r^*,\nu^*)$, where $r^* = t + r = \sum y_i + r, \nu^* = n + \nu$.
**Normal-Normal** (575) - $Y_i$ ~Normal$(\theta,\sigma_o^2)$ and $\pi$ ~Normal$(\mu_o,\tau_o^2)$ have posterior Normal $(\mu_o^*,\sigma_o^*)$, where $\mu_o^* = \frac{\mu_o\sigma_o^2 + n\tau_o^2\bar{Y}_n}{\sigma_o^2+n\tau_o^2}, \sigma_o^* = \frac{\sigma_o^2\tau_o^2}{\sigma_o^2+n\tau_o^2}$.
**Normal-Gamma** conjugate prior for $N(\theta,\sigma^2)$ with both unknown (578)
**Exponential family** conjugate priors (590)

## Loss Functions $L(\hat{\tau},\tau)$

**squared error loss function** (581) – $[\hat{\tau}-\tau(\theta)]^2$
**absolute error loss function** (581) – $|\hat{\tau}-\tau|$
**relative squared error loss function** (581) – $[\frac{\hat{\tau}-\tau}{\tau}]^2$
**Stein loss function** (581) – $\frac{\hat{\tau}-\tau}{\tau} - \log\frac{\hat{\tau}}{\tau} = \frac{\hat{\tau}-\tau}{\tau} - \log(1+\frac{\hat{\tau}-\tau}{\tau})$

# Chapter 8

# Hypothesis Tests

## 8.1 The Neyman-Pearson Lemma and UMP Tests

### Review

**null hypothesis** (598) – $H_o$
**alternative hypothesis** (598) – $H_a$
**decision rule** (598) – $\phi$ – a rule defining when to reject $H_o$ (e.g. Reject $H_o$ if $\mathsf{X} \in C$)
**test/critical function** (598) – $\phi : S \to [0, 1]$, where $\phi(\mathsf{x})$ is the conditional probability that $H_o$ is rejected given that $\mathsf{X}$ has the observed value $\mathsf{x}$.
**rejection region, critical region** $C$ (598) – typically consists of values $\mathsf{X}$ in the sample space for which some statistic $T$ takes on values implausible under $H_o$.
**Bayesian Approach:**
**risk** (604) – $R_\pi(\phi)$ is the expected loss associated with the decision rule $\phi$
**Bayes rule** (604) – for the prior $\pi(\cdot)$, any rule $\phi_\pi$ that minimizes the risk $R_\pi(\phi)$.
**Bayes risk** (604) – the risk as defined by such a rule $\phi_\pi$.
**conditional/posterior risk** (604) – $R_\pi(\phi|\mathsf{X}{=}\mathsf{x})$, the expected loss associated with the decision rule, given observations $\mathsf{x}$

### Definitions

**simple hypothesis** (598) – $H_o : \theta = \theta_o$ (one value)
**composite hypothesis** (598) – more than one value in $H_o(\underline{\Theta}_o)$
**level** (598) – aimed-for size, can be less than or equal to size; may not be achieved in discrete distributions
**size** (598) – $\alpha = \sup_{\underline{\Theta}_o} \beta(\theta)$ = actual level of the test
**randomized test** (598) – determined by test function, and used for discrete distributions to get level equal to $\alpha$
**power** (599) = $\beta(\theta) = E_\theta[P_\theta(H_o \text{ is rejected}|\mathsf{X})] = E_\theta[\phi(x)]$
**no data test** (599) – $\phi(\mathsf{X}) = \alpha$ for all $\mathsf{x}$, always rejecting at random
**non-randomized test** (599) – $\phi(\mathsf{X}) = 1_C(\mathsf{X})$ for some rejection set $C$.
**Neyman-Pearson likelihood ratio statistic** (LR-statistic) (599) – $\Lambda(\mathsf{x}) = f_a(\mathsf{x})/f_o(\mathsf{x})$

### Theorem 8.1.1 – Neyman-Pearson Lemma (599)

Fix $0 \le \alpha \le 1$. Suppose $P_o$ and $P_a$ have distrtibutions $f_o$ and $f_a$.
(i) There exists a constant $k$ and a function $\gamma(\mathsf{x})$ for which the test

$$(\mathbf{a})\ \phi(\mathsf{x}) = \begin{cases} 1 \\ \gamma(\mathsf{x}) & \text{if } \Lambda(\mathsf{x}) = \frac{f_a(\mathsf{x})}{f_o(\mathsf{x})} = \\ 0 \end{cases} \begin{cases} > k \\ = k \\ < k \end{cases} \text{ has } (\mathbf{b})\ E_o[\phi(\mathsf{X})] = \alpha.$$

(ii) Any test of this type is a most powerful test of $P_o$ vs. $P_a$ of level $\alpha$.

**Corollary 1**
If $\Lambda(\mathsf{x}) = f_a(\mathsf{x})/f_o(\mathsf{x}) = g(t)$ for some $\uparrow$ function $g(\cdot)$ of $t = T(\mathsf{x})$, then the test in (i) above may be replaced (for appropriate constants $c$ and $\gamma$ that do exist) by

$$(\mathbf{a})\ \phi(\mathsf{x}) = \begin{cases} 1 & \text{if } t > c \\ \gamma & \text{if } t = c \\ 0 & \text{if } t < c \end{cases} \quad \text{where } (\mathbf{b})\ P_o(T > c) + \gamma P_o(T = c) = \alpha.$$

### Theorem 8.1.2 – Uniformly Most Powerful (UMP) tests (601)

If $\phi_o(\cdot)$ is a fixed test that gives a most powerful test for the simple hypothesis $\theta_o$ versus the simple alternative $\theta_a$, and that this same $\phi_o$ test is most powerful against each fixed alternative $\theta_a \in \underline{\Theta}_a$, then
$\phi_o$ is a UMP test of $H_o : \theta_o$ vs $H_a : \theta \in \underline{\Theta}_a$.

If this same test is a level $\alpha$ test for the composite hypothesis $\theta \in \underline{\Theta}_o$ (that is $[\sup_{\theta \in \underline{\Theta}_o} \beta_{\phi_o}(\theta)] \leq \alpha$, then
$\phi_o$ is a UMP test of $H_o : \theta_o \in \underline{\Theta}_o$ vs $H_a : \theta \in \underline{\Theta}_a$.

### Role of sufficient statistics (601)

The Neyman-Pearson test may be based directly on any sufficient statistic rather than the full likelihood by application of the factorization theorem (that $f_\theta(\mathsf{x}) = g_\theta(t)h(\mathsf{x})$):
$\Lambda^{NP}(\mathsf{x}) = f_{\theta_a}(\mathsf{x})/f_{\theta_o}(\mathsf{x}) = g_{\theta_a}(t)/g_{\theta_o}(t) = \Lambda^T(t)$.

### Theorem 8.1.3 – Karlin-Rubin (601)

Define *monotone likelihood ratio* for a family of distributions $\{f_\theta : \theta \in \underline{\Theta}\}$ if for $\theta < \theta'$, $\frac{f_{\theta'}(\mathsf{x})}{f_\theta(\mathsf{x})}$ is an increasing (or decreasing) function of $t = T(\mathsf{x})$.

For a family with MLR in $T$, the test $\phi$ defined in the Neyman-Pearson Corollary is UMP for testing $H_o : \theta \leq \theta_o$ vs $H_a : \theta > \theta_o$.

### Exponential families with MLR (602)

Model the $\mathsf{X}$-experiment as having one of the distributions in the exponential family
$c(\theta)e^{\eta(\theta)T(\mathsf{X})}h(x)$, with natural parameter $\eta(\theta)$ that is $\uparrow$ in $\theta$.
Then the family of distributions $T$ has MLR in $T$. (Likewise if $\eta(\theta)$ is $\downarrow$ in $\theta$.

## 8.2 The Likelihood Ratio Test

### Definitions

**Likelihood Ratio Test** (610) – $\Lambda^{LR} = \dfrac{sup_{\underline{\Theta}} L(\mathbf{X}, \theta)}{sup \underline{\Theta}_o L(\mathbf{X}, \theta)} = \dfrac{L(\hat{\theta}_n; \mathsf{x})}{L(\hat{\theta}_n^o; \mathsf{x})}$

**Kullbak-Leibler information number** (626) – $H_{\theta,\theta_o} = K(P_a, P_o) = E_{P_a}[\log \frac{p_a(\mathsf{X})}{p_o(\mathsf{X})}] \geq 0$ compares $P_a$ and $P_o$, and is a well defined number for *every* possible pair $P_a$ and $P_o$ of probability distributions.

### Theorem 8.2.1 – Asymptotic null distribution of the LR-test statistic $\Lambda_n$ (610)

Let $r$ denote the dimension of the vector parameter $\theta = (\theta_1, \ldots, \theta_r)'$, and express $H_o : \theta_{q+1} = \cdots = \theta_r = 0$.

$2 \log \Lambda_n(\mathbf{X}) \rightarrow_d \chi^2_{r-q}$ when $H_o$ is true.

Asymptotically, the level $\alpha$ LR-test rejects $H_o$ in case $2 \log \Lambda_n$ exceeds $\chi^{2\,(\alpha)}_{r-q}$.

## Improvements on Asymptotic LR-test for Specific Cases (610)

(i) If $\Lambda_n$ is an $\uparrow$ function of some simpler statistic $T = T(\mathsf{X})$ whose distribution is exactly known under $H_o$, the LR-test rejects $H_o$ if $T$ is "too big".

(ii) If $\Lambda_n$ is a u-shaped function of such a statistic $T$, the LR-test rejects $H_o$ for both extremely large and extremely small values of $T$.

## Theorem 8.2.2 – Asymptotic behavior of $\lambda_n$ under alternatives (610)

Let $\underline{\Theta}_o = \theta_o$. Then

$$\frac{2}{n} \log \Lambda_n(\mathsf{X}) \to_p 2 \times (\text{some number } H_{\theta,\theta_o}) > 0 \text{ when } \theta \text{ is true.}$$

Thus, the power of the LR-test converges to 1 for every $\theta \in H_a$.

## Math Fact (expanding $\log(1+z)$)

$\log(1+z) = z - z^2/2 + z^2 g(z)$, where $g(z) \to 0$ as $z \to 0$.

## Examples

General result for Normal Theory Linear Model (614)
Theorem 8.2.3 – Asymptotic null distribution of various statistics (618)
- (a) Wald Test
- (b) Rao Score Test
- (c) LR-Test
- (d) Consistency of Observed Information
- (e) m-Dimensions

Maxwell($\theta$) practice (619)

# Chapter 9

# Regression, Anova, and Categorical Data

## 9.1 Simple Linear Regression

### Definitions

**simple linear regression model** (634) – $Y_i = \alpha + \beta(x_i - \bar{x}) + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma^2)$. An alternative parameterization: $Y_i = \gamma + \beta x_i + \epsilon_i$

$\alpha$ (634) – intercept of fitted regression line at $\bar{x} = 0$

$\gamma$ (634) – intercept of fitted regression line at $Y = 0$

$\beta$ (634) – slope; mean increase in the output per unit increase in the input $x$.

**least squares estimators** (LSEs) (635) – estimators $\hat{\alpha}, \hat{\beta}$ which minimize the sum of squares $\sum_{i=1}^{n}[\text{Observation}_i - \text{Expectation}_i]^2 = \sum_{i=1}^{n}[Y_i - (a + b(x_i - \bar{x}))]^2$. They solve the *normal equations.*

**normal equations** (634) – $0 = n[\bar{Y} - a]$; $0 = \sum_{i=1}^{n}(x_i - \bar{x})Y_i - b\sum_{i=1}^{n}(x_i - \bar{x})^2$

**fitted values** (635) – $\hat{Y}_i = \hat{\alpha} + \hat{\beta}(x_i - \bar{x})$

**residuals** (635) – $\hat{\epsilon}_i = Y_i - \hat{Y}_i$

**studentized residuals** (635) – $\hat{Z}_i = \hat{\epsilon}_i / S_n$; these form ancillary statistics

**coefficient of determination** (639) – $R^2 = SS_{xY}/\sqrt{SS_{xx}SS_{YY}} = \dfrac{SS_{YY} - SS_E}{SS_{YY}}$.

### Estimators **(635)**

The least squares estimators as defined above give:
$\hat{\alpha} = \bar{Y}$
$\hat{\beta} = \frac{1}{SS_{xx}}\sum_{i=1}^{n}(x_i - \bar{x})Y_i = \frac{1}{SS_{xx}}\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y}) = \frac{SS_{xY}}{SS_{xx}}$
$\hat{\gamma} = \bar{Y} - \hat{\beta}\bar{x}$

An estimate of $\sigma^2$ is:
$S_n^2 = \frac{1}{n-2}SS_E = \frac{1}{n-2}\sum_{i=1}^{n}\hat{\epsilon}_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}(x_i - \bar{x}))^2$

$\hat{\alpha}, \hat{\beta}$,and $S_n^2$ are independent, sufficient, complete, and independent of the studentized residuals (which are $\therefore$ ancillary).

### Behavior of $S_n^2$ **(637)**

$S_n^2 = \frac{1}{n-2}SS_E$ $\quad$ $E[S_n^2] = \sigma^2$ $\quad$ $S_n^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{\epsilon}_i^2 \sim \frac{1}{n-2}\sigma^2\chi_{n-2}^2$

This knowledge is useful in characterizing the behavior of $\hat{\alpha}$ and $\hat{\beta}$.

### Behavior of $\hat{\alpha}$ (635,637)

$\hat{\alpha} \sim (\alpha, \frac{1}{n}\sigma^2)$ $\qquad$ $\sqrt{n}(\hat{\alpha} - \alpha) = \sqrt{n}\bar{\epsilon} \sim N(0, \sigma^2)$ $\qquad$ $\dfrac{\hat{\alpha} - \alpha}{S_n/\sqrt{n}} \sim \mathcal{T}_{n-2}$

$\hat{\alpha} \pm t_{n-2}^{(\alpha/2)} \dfrac{S_n}{\sqrt{n}}$ yields a $1 - \alpha$ confidence interval for $\alpha$.
Reject $H_o : \alpha = \alpha_o$ vs $H_a : \alpha > \alpha_o$ when $T_n(\alpha_o) > t_{n-2}^{(\alpha)}$.

$$\text{Power}(\alpha, \sigma) = P_{\alpha,\sigma}\left(\frac{Z_\alpha + \delta_\alpha}{\sqrt{\chi^2_{n-2}/(n-2)}} > t_{n-2}^{(\alpha)}\right) = P(\mathcal{T}_{n-2}(\delta_\alpha) > t_{n-2}^{(\alpha)})$$

### Behavior of $\hat{\beta}$ (637)

$\hat{\beta} \sim (\beta, \sigma^2/SS_{xx})$ $\qquad$ $\sqrt{SS_{xx}}(\hat{\beta} - \beta) = \sum_{i=1}^n d_{ni}\epsilon_i \sim N(0, \sigma^2)$ $\qquad$ $\dfrac{\hat{\beta} - \beta}{S_n/\sqrt{SS_{xx}}} \sim \mathcal{T}_{n-2}$

$\hat{\beta} \pm t_{n-2}^{(\alpha/2)} \dfrac{S_n}{\sqrt{SS_{xx}}}$ yields a $1 - \alpha$ confidence interval for $\beta$.
Reject $H_o : \beta = \beta_o$ vs $H_a : \beta > \beta_o$ when $T_n(\beta_o) > t_{n-2}^{(\alpha)}$.

$$\text{Power}(\beta, \sigma) = P_{\beta,\sigma}\left(\frac{Z_\beta + \delta_\beta}{\sqrt{\chi^2_{n-2}/(n-2)}} > t_{n-2}^{(\alpha)}\right) = P(\mathcal{T}_{n-2}(\delta_\beta) > t_{n-2}^{(\alpha)})$$

### Extensions

**Behavior of $l(x_o) = \alpha + (x_o - \bar{x})\beta$ at another $x_o$** (638)
**Prediction Intervals** (638)
**Simultaneous Confidence Intervals** for $\alpha, \beta$, and the line $l(\cdot)$ (638)
**Exact Confidence Intervals** without the Normal RVs Assumption (639)
**Maximum Likelihood Estimates** (MLEs), Likelihood Ratio Tests (640)

## 9.2 One Way Analysis of Variance

## 9.3 Categorical Data

### Definitions

**observed cell counts** (654) – $(N_1, \ldots, N_\kappa)$
**expected cell counts** (654) – $E_1 = np_{o1}, \ldots, E_\kappa = np_{o\kappa}$
**normed differences** (ndiffs) (654) – $D_i = \text{NDiff}_i = \dfrac{N_i - E_i}{\sqrt{E_i}} = \dfrac{N_i - np_{oi}}{\sqrt{np_{oi}}} = \dfrac{\sqrt{n}(\hat{p}_i - p_{oi})}{\sqrt{p_{oi}}}$

### Theorem 9.3.1 – Chisquare tests of fit for Multinomial observations (655)

For $(N_1, \ldots, N_\kappa) \sim$Multinomial$(n; p_1, \ldots, p_\kappa)$, define the null hypothesis $H_o : p_1 = p_{o1}, \ldots, p_\kappa = p_{o\kappa}$

$$\hat{\chi}^2_n = \sum_{i=1}^{\kappa}\left[\frac{\sqrt{n}(\hat{p}_i - p_{oi})}{\sqrt{p_{oi}}}\right]^2 \to_d \chi^2_{\kappa-1} \qquad \text{when } H_o \text{ is true.}$$

$$\tfrac{1}{n}\hat{\chi}^2_n \to_p d_{\mathsf{p}} = \sum_{i=1}^{\kappa} \frac{(p_i - p_{oi})^2}{p_{oi}}$$

$$\hat{\chi}^2_n \to_p \infty \qquad \text{whenever } H_o \text{ is false.}$$

These results lead to a level-$\alpha$ test: Reject $H_o$ if $\chi^2_n > \chi^{2(\alpha)}_{\kappa-1}$ and asymptotic power function: Power$(\mathsf{p}) = P_{\mathsf{p}}(\chi^2_n > \chi^{2(\alpha)}_{\kappa-1}) \to_p 1$ as $n \to \infty$.

### Theorem 9.3.2 – Chisquare test for independence (659)

For the $IJ$ rvs $N_{ij}$ ~Multinomial($n;\mathbb{P}$), define the null hypothesis $H_o : \mathbb{P}_o$

$$\hat{\chi}_n^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(N_{ij}-N_{i\cdot}N_{\cdot j}/n)^2}{n_{i\cdot}N_{\cdot j}/n} \rightarrow_d \chi^2_{(I-1)(J-1)} \quad \text{when } H_o : \mathbb{P}_o \text{ is true.}$$

$$\tfrac{1}{n}\hat{\chi}_n^2 \rightarrow_p d_{\mathbb{P}} = \sum_{i=1}^{I}\sum_{j=1} J\frac{(p_{ij}-p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}}$$

$$\hat{\chi}_n^2 \rightarrow_p \infty \quad \text{whenever } H_o \text{ is false.}$$

These results lead to a level-$\alpha$ test: Reject $H_o$ if $\chi_n^2 > \chi^{2(\alpha)}_{(I-1)(J-1)}$.
and asymptotic power function: Power($\mathbb{P}$) $= P_{\mathbb{P}}(\hat{\chi}_n^2 > \chi^{2(\alpha)}_{(I-1)(J-1)}) \rightarrow_p 1$ as $n \rightarrow \infty$.

### Theorem 9.3.3 – Chisquare test for independence via the conditionals (663)

For $(N_{i1},\ldots,N_{iJ})$ ~Multinomial($n_{i\cdot};p_{1|i},\ldots,p_{J|i}$), define the null hypothesis $H_o : \mathbb{P}_o$

$$\hat{\chi}_n^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(N_{ij}-n_{i\cdot}N_{\cdot j}/n)^2}{n_{i\cdot}N_{\cdot j}/n} \rightarrow_d \chi^2_{(I-1)(J-1)} \quad \text{when } H_o : \mathbb{P}_o \text{ is true.}$$

$$\tfrac{1}{n}\hat{\chi}_n^2 \rightarrow_p d_{\mathbb{P}} = \sum_{i=1}^{I}\sum_{j=1} J\frac{n_{i\cdot}}{n}\frac{(p_{j|i}-\bar{p}_j)^2}{\bar{p}_j}$$

$$\hat{\chi}_n^2 \rightarrow_p \infty \quad \text{whenever } H_o \text{ is false.}$$

These results lead to a level-$\alpha$ test: Reject $H_o$ if $\chi_n^2 > \chi^{2(\alpha)}_{(I-1)(J-1)}$.
and asymptotic power function: Power($\mathbb{P}$) $= P_{\mathbb{P}}(\hat{\chi}_n^2 > \chi^{2(\alpha)}_{(I-1)(J-1)}) \rightarrow_p 1$ as $n \rightarrow \infty$.

### Theorem 9.3.4 – Chisquare test for independence with marginals fixed (665)

$$\hat{\chi}_n^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(N_{ij}-n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n} \rightarrow_d \chi^2_{(I-1)(J-1)} \quad \text{when } H_o \text{ is true.}$$

A level-$\alpha$ test is: Reject $H_o$ if $\chi_n^2 > \chi^{2(\alpha)}_{(I-1)(J-1)}$.

### Degrees of Freedom Rule (666)

When $H_o$ is true,
$df$ =[number of cells]-[number of parameters estimated]-[number of cell count restrictions].

| | |
|---|---|
| Chisquare test of fit: | $df = [\kappa]-[0]-[1] = \kappa-1$ |
| Testing for independence: | $df = [IJ]-[(I-1)+(J-1)]-[1] = (I-1)(J-1)$ |
| Testing for independence via the conditional distributions: | $df = [IJ]-[(J-1)]-[I] = (I-1)(J-1)$ |
| Testing for independence with all marginals fixed: | $df = [IJ]-[0]-[I+J-1] = (I-1)(J-1)$ |

### Proofs that $\hat{p}_i$ is the MLE and unique UMVUE of $p_i$ (656)

Lagrange Multipliers (656)
Jensen's Inequality (671)
Guess and verify (671)
Maximizing a function of $\kappa - 1$ variables (671)

# Appendix A

# Distributions

## A.1 Important Transformations

**Cauchy(0,1)** **(142)**

$f(x) = \dfrac{1}{\pi(1+x^2)}$ on $-\infty < x < \infty$

**Cauchy$(\alpha, \beta)$** **(373)**

$f_{\alpha,\beta}(x) = \dfrac{1}{\pi\beta}\dfrac{1}{1+((x-\alpha)/\beta)^2}$ on $-\infty < x < \infty$

**Inverted Gamma$(r, \theta)$** **(142)**

Let $M = 1/W$, where $W$ has the Gamma$(r, \theta)$ distribution.
Then $M$ has the Inverted Gamma$(r, \theta)$ distribution:
$f_M(m) = -\dfrac{1}{\Gamma(r)}\dfrac{1}{\theta^r}\dfrac{1}{m^{r+1}}e^{-1/m\theta}$

**Rayleigh$(\theta)$ distribution** **(417)**

$f_\theta(x) = \frac{x}{\theta^2}e^{-\frac{1}{2}x^2/\theta^2}$ for $x \geq 0$

**Maxwell$(\theta)$ distribution** **(417)**

$f_\theta(x) = \sqrt{\frac{2}{\pi}}\frac{x^2}{\theta^3}e^{-\frac{1}{2}x^2/\theta^2}$ for $x \geq 0$

**Exponential$(\alpha, \nu)$** **(373)**

$f_{\alpha,\mu}(y) =^{-nu(y-\alpha)}$ on $\alpha \leq y < \infty$

**Logistic(0,1) distribution** **(426)**

$f(x) = e^{-x}/(1+e^{-x})^2$ on $-\infty < x < \infty$

**Logistic$(\alpha, \beta)$** **(373)**

$f_{\alpha,\beta}(x) = \dfrac{\exp(-(x-\alpha)/\beta)}{\beta[1+\exp(-(x-\alpha)/\beta)]^2}$ on $-\infty < x < \infty$

**Double Exponential**$(\mu, \theta)$ **(33)**

$f_X(x) = \frac{1}{2\theta}\exp(-|x-\mu|/\theta)$ on $-\infty < x < \infty$

**Pareto**$_2(1, \nu)$ **(531)**

$f_\nu(x) = \nu/(1+x)^{1+\nu}$ for $x > 0$

**Truncation families (490,509,521)**

$f_\theta(x) = c(\theta)1_{[0,\theta]}(x)h(x)$ for $x \geq 0$

**Power family (509,531)**

$f_{\beta,\gamma}(x) = \frac{\gamma}{\beta}(\frac{x}{\beta})^{\gamma-1}1_{[0,\beta]}(x)$

$f_\gamma(x) = \dfrac{\gamma-1}{x^\gamma}$ for $x > 1$

**Chisquare (140)**

$f_k(y) = \dfrac{y^{\frac{k}{2}-1}}{\Gamma(\frac{k}{2})2^{\frac{k}{2}}}e^{-y/2}$ for all $y \geq 0$.

**Logarithmic Series (144,530)**

$p_\theta(k) = \dfrac{(1-\theta)^k}{(log1/\theta)k}$ for $k = 1, 2, \ldots$

## Additional Distributions

**Uniform(a,b) (33)**

**Discrete Uniform (1,N) (33)**

**Discrete Uniform (0,N) (33)**

**Triangular(c,a) (33)**

**Student Correlation Coefficient (145)**

**Empirical (146)**

**Truncated Poisson (301)**

**Generalized Binomial (319)**

**Studentized range distribution (371)**

**LogNormal**$(\mu, \sigma^2)$ **(373)**

**LogGamma(**$\kappa$**,0,1) distribution (374,427)**

**LogGamma(1,0,1)=ExtValueMin(0,1)=Gompertz distribution (374,427,429)**

**Multivariate Normal**$(\mu, \Sigma)$ **distribution (385)**

**Slash distribution (403)**

**Inverted Gamma(a,b) (577)**

**Dirichlet distribution (587)**