# 1 Distributions

## 1.1 Binomial

Sum of $n$ bernoulli trials with probability of success $p$.

$X \sim B(n,p)$

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n$$

mgf: $M_X(t) = (pe^t + 1 - p)^n$

$\mathrm{E}\, X = np, \quad \mathrm{Var}\, X = np(1-p)$

## 1.2 Poisson

$X \sim P(\lambda)$

$$p(k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad k = 0, 1, \ldots$$

mgf: $M_X(t) = e^{\lambda(e^t - 1)}$

$\mathrm{E}\, X = \lambda, \quad \mathrm{Var}\, X = \lambda$

## 1.3 Multivariate Normal

$X \sim N(\mu, \Sigma)$, $\Sigma$ positive definite

$$f(x) = \frac{\exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\}}{(2\pi)^{\frac{k}{2}}\sqrt{\det(\Sigma)}}$$

## 1.4 Beta

$X \sim \mathrm{Beta}(\alpha, \beta)$

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \qquad 0 \le x \le 1$$

using the beta function:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$$

$\mathrm{E}\, X = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}\, X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

## 1.5 Gamma

$X \sim \mathrm{Gamma}(\alpha, \lambda)$

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \qquad x > 0$$

mgf: $M_X(t) = (1 - \frac{t}{\lambda})^{-\alpha}, t < \lambda$

$X \sim \mathrm{Gamma}(\alpha, \lambda) \iff \lambda X \sim \mathrm{Gamma}(\alpha, 1)$

$X_i$ iid $\mathrm{Gamma}(\alpha_i, \lambda)$, then

$$\sum X_i \sim \mathrm{Gamma}(\sum \alpha_i, \lambda)$$

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

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

$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

$\Gamma(k) = (k-1)!$ for $k$ positive integer.

## 1.6 Exponential

Special case: $\mathrm{Exp}(\lambda) \equiv \mathrm{Gamma}(1, \lambda)$

## 1.7 Chi square

Special case: $\chi^2_n \equiv \mathrm{Gamma}(\frac{n}{2}, \frac{1}{2})$

Let $Z_i$ be iid $N(0,1)$.

$\sum_{i=1}^n Z_i^2 \sim \chi^2_n$

Noncentral $\chi^2$. Let $Y \sim N(\mu, I)$ be an $n$ vector. Then

$$\|Y\|^2 \sim \chi^2_n(\|\mu\|^2)$$

## 1.8 F

$$F(m,n) \equiv \frac{\frac{\chi^2_m}{m}}{\frac{\chi^2_n}{n}}$$

Where numerator and denominator are independent $\chi^2$.

## 1.9 T

$$t(n) = \frac{N(0,1)}{\sqrt{\frac{\chi^2_n}{n}}}$$

Where numerator and denominator are independent.

### 1.10 Studentized range distribution

Let $z_1, \dots z_K$ be i.i.d $N(0,1)$ and $X$ be an independent $\chi^2_\nu$. Then

$$\frac{\max_i z_i - \min_i z_i}{\sqrt{(}X/\nu)} \sim q_{K,\nu}$$

# 2 Math

Moment generating functions determine distribution

$$M_X(t) \equiv \mathrm{E}(e^{tX}), \quad M'_X(0) = \mathrm{E}(X)$$

For $X_i$ independently distributed:

$$M_{\sum X_i}(t) = \prod M_{X_i}(t)$$

Use this to find distributions of sums of RV's.

## 2.1 Inequalities

**Jensen's Inequality** if $S \subset R^k$ convex and closed, $g$ convex on $S$, $P[X \in S] = 1$, and $\mathrm{E}\, X$ is finite, then $\mathrm{E}\, X \in S$, $\mathrm{E}\, g(X)$ exists, and

$$\mathrm{E}\, g(X) \geq g(\mathrm{E}\, X)$$

**Holder's Inequality** if $r, s > 1$ and $\frac{1}{r} + \frac{1}{s} = 1$ then

$$\mathrm{E}\,|XY| \leq (\mathrm{E}\,|X|^r)^{\frac{1}{r}} (\mathrm{E}\,|X|^s)^{\frac{1}{s}}$$

Order statistics for sorted sample $X_{(1)}, \dots, X_{(n)}$ has pdf:

$$n! \prod_{i=1}^{n} f(X_{(i)}) \quad I(X_{(1)} < \cdots < X_{(n)})$$

$T(X)$ Sufficient means the distribution of $X|T(X)$ does not depend on $\theta$.

Factorization theorem: $T(x)$ is sufficient $\iff$

$$f_\theta(x) = h(x) g(\theta, T(x))$$

$L_x(\theta) = p_\theta(x) = p(x, \theta)$ likelihood is function of $\theta$, density is function of $x$.

The likelihood ratio

$$\lambda_x(\theta) = \frac{L_x(\theta)}{L_x(\theta_0)}$$

is minimal sufficient. To show $T(x)$ is minimal sufficient show that it is sufficient and a function of the likelihood $\lambda_x(\theta)$.

**Fisher information**

$$I(\theta) = \mathrm{E}_\theta \left[ \frac{\partial}{\partial \theta} \log L_X(\theta) \right]^2 = \mathrm{E}_\theta \left[ -\frac{\partial^2}{\partial \theta^2} \log L_X(\theta) \right]$$

## 2.2 Bias, Variance

bias $\hat{v} \equiv \mathrm{E}(\hat{v}) - v$

$$MSE(\hat{v}) \equiv \mathrm{E}(\hat{v} - v)^2 = \mathrm{Var}(\hat{v}) + (\text{bias } \hat{v})^2$$

**Rao-Blackwell** Let $S(X)$ be an unbiased point estimator for $g(\theta)$. Conditioning on a sufficient statistic $T(X)$ reduces variance.

$$\mathrm{Var}_\theta(S(X)) \geq \mathrm{Var}_\theta(\mathrm{E}(S(X)|T(X)))$$

Also holds for more general convex loss function $L$:

$$R(\theta, S) \equiv \mathrm{E}_\theta\, L(\theta, S(X)) \geq \mathrm{E}_\theta\, L(\theta, \mathrm{E}(S(X)|T(X)))$$

**Completeness** $T(X)$ is complete if $\mathrm{E}\, g(T(X)) = 0$ implies $g = 0$ almost surely for all $\theta$.

Basu's Theorem - If $T(X)$ complete sufficient statistic and $A(X)$ is ancillary then $A(X)$ and $T(X)$ are independent.

**Lehmann - Scheffe** Suppose $T(X)$ is complete sufficient. Then there exists unique unbiased estimator $\mathrm{E}\, h(T(x))$ of $g(\theta) \in R$ with smallest variance (MVUE)

## 2.3 Exponential Families

One parameter model indexed by $\theta$.

$$p(x, \theta) = h(x) \exp\{\eta(\theta) T(x) - B(\theta)\}$$

Canonical form model indexed by $\eta$.

$$q(x, \eta) = h(x) \exp\{\eta T(x) - A(\eta)\}$$

Then moment generating function for $T(X)$ is

$$M_{T(X)}(t) = \exp\{A(t + \eta) - A(\eta)\}$$

## 2.4 Multivariate Normal

Stein's formula: $X \sim N(\mu, \sigma)$

$$\mathrm{E}(g(X)(X - \mu)) = \sigma^2\, \mathrm{E}(g'(X))$$

assuming these expectations are finite.

$X \sim N(\mu, \Sigma)$, $A$ an $m \times n$ matrix, then

$$AX \sim N(A\mu, A\Sigma A^t)$$

For $\Sigma$ full rank it's possible to transform between $z \sim N(0, I)$ and $X$:

$$X = \Sigma^{1/2} z + \mu \qquad z = \Sigma^{-1/2}(X - \mu)$$

Conditional distribution:

## 2.5 Conditional Distributions

Conditional pdf:

$$f_{x|y}(x, y) \equiv \frac{f_{x,y}(x, y)}{f_y(y)}$$

Iterated expectation:

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

Conditional variance formula:

$$\mathrm{Var}(Y) = \mathrm{Var}(E(Y|X)) + E(\mathrm{Var}(Y|X))$$

# 3 Applied

Matrix / Vector differentiation

$\frac{\partial A^T\beta}{\partial\beta} = A$, $\frac{\partial \beta^T A\beta}{\partial\beta} = (A + A^t)\beta = 2A\beta$ for $A$ symmetric.

## 3.1 Linear Models

Least Squares Principle

$$\arg\min_\beta \|Y - X\beta\|^2$$

Normal Model

$$Y = X\beta + \epsilon, \qquad \epsilon \sim N(0, \sigma^2 I)$$

Normal Equations - Any $b$ satistfying this solves the least squares

$$X^T X b = X^T y$$

Gauss Markov Theorem - $\hat{\beta}$ is Best Linear Unbiased Estimator (BLUE) of $\beta$.

$$\hat{\beta} = (X^T X)^{-1} X^T y \sim N(\beta, \sigma(X^T X)^{-1})$$

Estimating the variance: $\frac{\|y - X\hat{\beta}\|^2}{\sigma^2} \sim \chi^2_{n-p}$.

$$\hat{\sigma}^2 = \frac{\left\|y - X\hat{\beta}\right\|^2}{n - p}$$

Use t test for hypothesis testing and confidence intervals for the value of a particular $\beta_j$ coefficient. Let $w_{ii}$ be the $i$th diagonal entry of $(X^T X)^{-1}$.

$$\frac{\beta_j - \beta_j^*}{\hat{\sigma}\sqrt{w_{ii}}} \sim t_{n-p}$$

$1 - \alpha$ Confidence intervals for new observation $Y_h$ at $x_h$ and $E[Y_h]$:

$$E[y_h] \approx \hat{y_h} \pm t(n-p, 1 - \frac{\alpha}{2})\hat{\sigma}\sqrt{x_h^T (X^T X)^{-1} x_h}$$

$$y_h \approx \hat{y_h} \pm t(n-p, 1 - \frac{\alpha}{2})\hat{\sigma}\sqrt{1 + x_h^T (X^T X)^{-1} x_h}$$

General linear tests. Partition $\beta = (\beta_1, \beta_2)$ where $\beta_1$ is an $r$ vector and $\beta_2$ is $p - r$. Null hypothesis $H_0 : \beta_2 = \beta_2^*$ (often 0), and $H_a : \beta_2 \neq \beta_2^*$. Then $SSE_r = \left\|y - X_2\beta_2^* - X_1\tilde{\beta}_1\right\|^2$ is the sum of squared error for the reduced model and $SSE_f = \left\|y - X\hat{\beta}\right\|^2$ is the squared sum of error for the full model. Under $H_0$:

$$\frac{\frac{SSE_r - SSE_f}{p-r}}{\frac{SSE_f}{n-p}} \sim F_{p-r, n-p}$$

Alternate forms of linear test, and testing a linear combination if $R\beta = r$.

$$\frac{(R\hat{\beta} - r)^T (R(X^T X)^{-1} R^T)^{-1} (R\beta - r)/s^2}{\hat{\sigma}} \sim F_{s, n-p}$$

Simultaneous confidence bands for all $f(x)$.

If the model has an intercept then $SSTO = SSR + SSE$.

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i = \bar{y})^2 + (yi - \hat{yi})^2$$

$R^2$ and adjusted $R^2$

## 3.2 ANOVA

Three principles of experimental design: 1) Replication 2) Randomization 3) Blocking

One way ANOVA with $n$ total observations, $K$ groups:

| SS | | DF |
|---|---|---|
| SSTR | $\sum_{j=1}^K n_j(\bar{y_{j\cdot}} - \bar{y_{\cdot\cdot}})^2$ | K - 1 |
| SSE | $\sum_{i=1}^n (y_{ij} - \bar{y_{j\cdot}})^2$ | n - K |
| SSTO | $\sum_{i=1}^n (y_{ij} - \bar{y_{\cdot\cdot}})^2$ | n - 1 |

Contrasts are sums of the form $\Phi = \sum_{i=1}^K c_i\mu_i$ with $\sum_{i=1}^K c_i = 0$. Tukey's works for all pairwise contrasts. Scheffe's and extended Tukey works for all contrasts. Bonferroni's is for a limited number of pre specified contrasts.

**Ridge Regression** for $\lambda > 0$ solves

$$\min_\beta ||Y - X\beta||^2 + \lambda||\beta||^2$$

## 3.3 General Techniques

Singular Value Decompostion (SVD) Any matrix $X$ can be written

$$X = UDV^T$$

with $U, V$ orthogonal, and $D$ diagonal.

Moore Penrose Psuedoinverse $A^+$ exists uniquely for every matrix $A$.

Projection matrix $P$ are symmetric and idempotent. They have eigenvalues either 0 or 1.

$$P = P^T \qquad P^2 = P$$

Covariance of linear transformations

$$Cov(Ay, Bx) = ACov(y, x)B^T$$

# 4 Miscellaneous

Integration by parts:

$$\int uv' = uv - \int u'v$$

# 5 Course Specific

## 5.1 Applied Lectures

23 Nov - BIC, AIC derivations, Mallow's cp, stepwise selection algorithms, outliers and studentized residuals

18 Nov - F test with orthogonalized X, model selection criteria, bootstrap t method in lab

16 Nov - Multicollinearity, Variance Inflation Factor, ridge regression, bias variance tradeoff, AIC, BIC, cross validation, proof leave one out cross validation formula for OLS

11 Nov - Holiday

9 Nov - Linear model with random $X$, transformations of $y$ and $X$, box cox procedure, bootstrap with percentile-t and fixed $X$ sampling, weighted least squares

4 Nov - Interaction plots for two way ANOVA with balanced design, Linear models with random $X$

2 Nov - Midterm

28 Oct - Kronecker product formulae for two way ANOVA

26 Oct - Kronecker product 1 way ANOVA, decomposition of two way ANOVA, noncentral $\chi^2$ distributions for ANOVA table SSA, SSB, SSAB

21 Oct - Tukey's method for pairwise contrasts, Bonferroni's method, definition and properties of Kronecker product

## 5.2 Math Lectures

23 Nov - Fisher information, Cramer Rao inequality, Exponential families and properties

18 Nov - Rao-Blackwell theorem, Lehmann-Scheffe theorem, UMVUE examples for normal, uniform, Poisson, Fisher information

16 Nov - Minimal sufficiency, likelihood ratio, ancillary statistics, completeness, Basu's theorem, loss functions

11 Nov - Holiday

9 Nov - Distribution of order statistics, factorization theorem, sufficient statistics for Exponential families and uniform dist

4 Nov - Midterm

2 Nov - Location-scale families, invariance, ancillary and sufficient statistics, order statistics, multinomial distribution

28 Oct - Transformation of discrete and continuous random variables, Jacobian, examples with beta distributions, Dirichlet distribution

26 Oct - Jensen's and Holder's inequality, convex functions and sets, products of normal random variables

21 Oct - Convolution formula, examples with Uniform, Gamma, Poisson, marginal and conditional distributions for multivariate normal

Table 1: Problems in past 231 exams - Came from a brief glance at the question statements. TODO-make second updated table after solving questions that shows which techniques are used.

| | |
|---|---|
| Binomial | ******* |
| Poisson | ****** |
| Uniform | ****** |
| Normal | ****** |
| Gamma | *** |
| Exponential | ** |
| Negative binomial | * |
| Beta | * |
| Geometric | * |
| MLE | ************* |
| asymptotic distribution | *********** |
| Bayes estimator / risk | ********** |
| UMVUE / Cramer-Rao | ********* |
| minimiax | ****** |
| UMP test | ****** |
| linear regression | ***** |
| likelihood ratio | **** |
| Wald's test | *** |
| sufficient statistic | ** |
| Hierarchical model | * |
| method of moments | * |
| hypothesis testing | * |
| order statistics | * |

## 5.3 Problem Solving Strategies

**Read the whole question**

Read carefully and do the right problem! If there's a hint, it should probably be used. Early parts of a question can help for later parts, and later parts occasionaly provide insight for earlier parts.

**First principles**

When in doubt, work from definitions

**Look for Distributions**

Can the question be solved by knowing the distribution of some quantity? Example: $\sum(x_i - \bar{x})^2$ is $\chi^2_{n-1}$ for normal $x_i$.

**Fast and correct algebra**

Better to write more than to make a simple algebra mistake. Practice common manipulations so don't have to think about them when testing.

### 5.4 Math questions

1. State and prove the Lehmann Scheffe Theorem.

### 5.5 Applied questions

1. Derive ridge regression estimates, along with their bias and variances.