Matrix properties:

$A = \begin{bmatrix} 2 & 5 \\ 3 & 10 \\ 3 & 4 \end{bmatrix}$ $A \neq B$ (not equal)

$A' = \begin{bmatrix} 2 & 3 & 3 \\ 5 & 10 & 4 \end{bmatrix}$

$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$

- Symmetry: $A = A'$
- Diagonality: off-diagonal elements are zero
- Identity: Diagonal elements 1, others zero
- Scalar matrix: Diagonal elements some scalar, others zero
- $1_{n\times 1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ so $1'1 = [n]$, $11' = J_{n\times n}$
- $J_{r\times r} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$
- Rank = # linearly ind. columns
  ↳ can never exceed smaller rank when two matrices multiplied

Rules for matrix ranks:
- rank(A) = 0 iff A = 0 matrix
- if $A \neq 0$, $1 \leq \text{rank}(A) \leq \min(r,c)$
- rank(A) = rank(A')
- rank(A) = rank(A'A) = rank(AA')
- rank(A+B) ≤ rank(A) + rank(B)
- rank(AB) ≤ min[rank(A), rank(B)]

- $A^{-1}A = AA^{-1} = I$
- Idempotency: $AA = A$

Variance-stabilizing trans of Y → if var ↑ with $\hat{Y}$, go down the ladder and vice versa

Multicollinearity
- Two or more predictors highly correlated w/ each other
  ↳ If perfectly uncorrelated, marginal and partial relationships w/ outcome variable will be the same
- Effects of multicollinearity:

predictors	effect on $\hat{\beta}$	effect on SE($\hat{\beta}$)
uncorrelated	unchanged	unchanged or reduced
highly corr.	change, can be large	Increase, potentially large one

- What's going on w/ coefficient estimates?
  - In an MLR model, they are partial coefficients (holding other predictors constant / controlling)
  - In a SLR model, they are marginal coefficients (ignoring any other potential predictors)
  - In MLR, marginals and partials can even have opposite signs because you may be holding constant a predictor that naturally correlates w/ the one in question (i.e. BSA and weight). Model may not have much info about that rare situation.
- What's going on with SEs?
  - Intuitively, if X_1 and X_2 are highly correlated and we hold one constant, there is little variation in the other. Thus, the data contain little information about the partial effect of one on the outcome variable holding the other constant. The SEs of the partial coefs will be large, reflecting that there is little information about the situation in the data.
  - Mathematically, if X is not full rank, X'X isn't either, meaning $\det(X'X) = 0$, so inverse does not exist. If predictors close to linear dependence, $\det(X'X)$ close to 0 and $1/\det(X'X)$ large. Since $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$, the SEs will be high
- coef estimates, SE and pvalues can all be dramatically different when multicollinearity is present.

$Y'Y_{1\times 1} = [\sum Y_i^2]$

$X'X_{2\times 2} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$

$X'Y_{2\times 1} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$

- Inverse exists when:
  ↳ full rank, $\det \neq 0$

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

$(X'X)^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i-\bar{X})^2} & \frac{-\bar{X}}{\sum(X_i-\bar{X})^2} \\ \frac{-\bar{X}}{\sum(X_i-\bar{X})^2} & \frac{1}{\sum(X_i-\bar{X})^2} \end{bmatrix}$

Basic Results ↳ $(X'X)^{-1}$ exists iff rank(X) = p (p = # columns)

$A+B = B+A$
$(A+B)+C = A+(B+C)$
$(AB)C = A(BC)$
$C(A+B) = CA+CB$
$k(A+B) = kA+kB$
$(A')' = A$
$(A+B)' = A'+B'$
$(AB)' = B'A'$
$(ABC)' = C'B'A'$
$(AB)^{-1} = B^{-1}A^{-1}$
$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$
$(A^{-1})^{-1} = A$
$(A')^{-1} = (A^{-1})'$

Hat matrix

$H_{n\times n} = X(X'X)^{-1}X'$

$HH = H$ (idempotent)
$H' = H$ (symmetric)

$SSTO = Y'Y - (\frac{1}{n})Y'JY$
$SSE = e'e = (Y-X\hat{\beta})'(Y-X\hat{\beta}) = Y'Y - \hat{\beta}'X'Y$
$SSR = \hat{\beta}'X'Y - (\frac{1}{n})Y'JY$
Also, $\hat{\beta}'X' = (X\hat{\beta})' = \hat{Y}' = (HY)' = Y'H$

More ANOVA:
$SSTO = Y'[I - (\frac{1}{n})J]Y$
$SSE = Y'(I-H)Y$
$SSR = Y'[H - (\frac{1}{n})J]Y$

General MLR

$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}$; $\sigma^2\{\varepsilon\} = \sigma^2 I$

$E\{Y\} = X\beta$; $\sigma^2\{Y\} = \sigma^2 I$

$X'X\hat{\beta} = X'Y$
$\hat{\beta} = (X'X)^{-1}(X'X)Y$
$SSTO = Y'Y - (\frac{1}{n})Y'JY = Y'[I - (\frac{1}{n})J]Y$
$SSE = e'e = (Y-X\hat{\beta})'(Y-X\hat{\beta}) = Y'Y - \hat{\beta}'X'Y = Y'(I-H)Y$
$SSR = \hat{\beta}'X'Y - (\frac{1}{n})Y'JY = Y'[H - (\frac{1}{n})J]Y$

$\hat{Y} = X\hat{\beta}$; $e = Y - \hat{Y} = Y - X\hat{\beta}$
$H = X(X'X)^{-1}X'$
$e = (I-H)Y$
$\sigma^2\{e\} = \sigma^2(I-H)$
$s^2\{e\} = MSE(I-H)$

Diagnosing Multicollinearity:
- Pairwise correlation not sufficient
- Need to consider degree of linear relationship with the full set of other predictors in the model
- Use the R^2 obtained from regressing x_j on the other predictors (R_j^2)
- Recall $\text{var}(\hat{\beta}_j) = \frac{\sigma^2}{(n-1)s_j^2}$ when model only includes x_j.
- Now, with other predictors in the model, $\text{Var}(\hat{\beta}_j) = \frac{1}{1-R_j^2}\frac{\sigma^2}{(n-1)s_j^2}$
- $\frac{1}{1-R_j^2}$ = Variance Inflation Factor (VIF) for predictor j.
- When R_j^2 close to zero, VIF_j close to 1 and SE not much affected, and vice versa
- $\sqrt{\text{VIF}_j}$ gives impact on $SE(\hat{\beta}_j)$. VIF > 10 a problem, possibly smaller if interacting w/ others

- Two considerations for SE: $\text{Var}(\hat{\beta}_j) = \frac{1}{1-R_j^2}\frac{\hat{\sigma}^2}{(n-1)s_j^2}$ ($n-p$ effect on SE)
- Centering variables before creating interaction terms and polynomial terms can help reduce correlation and improve interpretation.
- Handling naturally correlated predictors
  ↳ if not interfering w/ important inferences, do nothing
  ↳ Consider changing predictors (often several variables represent same underlying construct)
  ↳ Often, when trying to explain or control for confounding, we can just choose one and not the other.

SLR

$Y_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$ $X_{n\times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$ $\beta_{2\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ $\varepsilon_{n\times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

$Y_{n\times 1} = X_{n\times 2}\beta_{2\times 1} + \varepsilon_{n\times 1}$

*Note $E\{Y\}_{n\times 1} = X_{n\times 2}\beta_{2\times 1}$

*Note assumption:

$\sigma^2\{\varepsilon\} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 I$

Normal equations: $X'X\hat{\beta} = X'Y$

To obtain the estimated regression coefficients: $(X'X)^{-1}X'X\hat{\beta} = (X'X)^{-1}X'Y$

$\hat{\beta} = (X'X)^{-1}X'Y$

Fitted: $\hat{Y} = X\hat{\beta}$ or $\hat{Y} = X(X'X)^{-1}X'Y = HY$

$\hat{\varepsilon} = Y - \hat{Y} = Y - X\hat{\beta} = Y - HY = Y(I-H)$ (residuals: estimates of the error terms)

$\sigma^2\{e\} = \sigma^2(I-H)$ and is estimated by $s^2\{e\} = MSE(I-H)$

Because $\sigma^2\{e\} = (I-H)\sigma^2\{Y\}(I-H)' = (I-H)\sigma^2 I(I-H) = \sigma^2(I-H)(I-H) = \sigma^2(I-H)$

$\sigma^2\{\hat{\beta}\} = \sigma^2(X'X)^{-1}$
$s^2\{\hat{\beta}\} = MSE(X'X)^{-1}$

Estimating mean response at X_h:

$X_h = \begin{bmatrix} 1 \\ X_h \end{bmatrix}$ (2×1)

So $\hat{Y}_h = X_h'\hat{\beta}$

$\sigma^2\{\hat{Y}_h\} = \sigma^2 X_h'(X'X)^{-1}X_h$
$= X_h'\sigma^2\{\hat{\beta}\}X_h$

$MSR = \frac{SSR}{p-1}$

$MSE = \frac{SSE}{n-p}$

... allow effect of x_1 on mean of y to change depending on value of x_2 with a product term:

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i$

- Relationship [illegible] when you group like terms
  - → β_1 is slope in x_{i1} when $x_{i2} = 0$
  - ↳ β_2 is slope in x_{i2} when $x_{i1} = 0$
  - ↳ β_3 is change in slope associated w/ a one-unit increase in the other predictor
- Center predictors if need be (for interp purposes)

Leverage:

- x_i is i^{th} row of design matrix $\quad H = X(X'X)^{-1}X'$
- Elements of H are h_{ij} (row, column) $\quad X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}$
  - ↳ $h_{ij} = x_i'(X'X)^{-1}x_j = x_j'(X'X)^{-1}x_i = h_{ji}$
  - ↳ $h_{ii} = x_i'(X'X)^{-1}x_i$
- Imagine SLR, where $h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$
  - ↳ So leverage is a type of relative distance measure
  - ↳ Leverage for obs i is proportion of the total squared deviations of the x_i from their mean $\bar{x}$ that is contributed by the observation (plus a $\frac{1}{n}$ term same for all obs.)
    - Far from mean, high leverage
    - close to mean, low leverage

Generally, $\bar{x}' = [\bar{x}_1, \ldots, \bar{x}_{p-1}]$ is mean vector, or centroid of predictor variables. $x_i^{*\prime} = x_i' - \bar{x}'$ is mean-centered version of covariate row vector for observation i

$$X^* = \begin{bmatrix} x_1^{*\prime} \\ \vdots \\ x_n^{*\prime} \end{bmatrix} = \begin{bmatrix} (x_{11} - \bar{x}_1) & \cdots & (x_{1,p-1} - \bar{x}_{p-1}) \\ \vdots & & \vdots \\ (x_{n1} - \bar{x}_1) & \cdots & (x_{n,p-1} - \bar{x}_{p-1}) \end{bmatrix}$$

$\frac{1}{n-1} X^{*\prime} X^*$ is sample variance-covariance matrix of predictors

$h_i = \frac{1}{n} + x_i^{*\prime}(X^{*\prime}X^*)^{-1}x_i^*$, so h_i measures distance from centroid taking into account covariance structure of X's

Partial regression plots (added-variable) are plots of residuals from partial regressions

Example plot of $e_{Y|x_2,\ldots,x_{p-1}}$ versus $e_{x_1|x_2,\ldots,x_{p-1}}$, along with partial regression line (from SLR regressing residuals on residuals)

- Helps with identifying influential obs., checking for non-linearity, suggesting monotonicity or not.

- Step functions
  - Break X into bins
  - Create cutpoints $c_1, \ldots, c_K$ called knots
    - $K+1$ indicator variables: $C_0(X) = I(X < c_1)$, $C_1(X) = I(c_1 \le X < c_2) \ldots$
    - ↳ piecewise constant function
- Piecewise linear reg:

Range X	Model
$x_i \le c_1$	$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
$c_1 < x_i \le c_2$	$Y_i = \beta_0 + \beta_1 x_i + \beta_2(x_i - c_1) + \varepsilon_i$
$\vdots$	$\vdots$

continuous at knots

- Regression spline
  - ↳ piecewise polynomial constrained to be smooth and cont. at knots.

$\beta_{10}C_1(x_i) + \beta_{11}C_1(x_i)x_i + \beta_{12}C_1(x_i)x_i^2 + \ldots$

not continuous at knots

Dropping β_{i0} after first terms (intercepts) gives continuity, but not smooth

Dropping second and third linear terms gives equal first derivs (duh)

Dropping quadratic terms gives equal curvature

Total diagnostics

- Conditional y-outliers vs. marginal y-outliers
  - ↳ If grouped w/ rest of y data, but is outlier in conditional dist. of Y given x_1, we have conditional y-outlier. } influential
  - ↳ If not grouped w/ rest of y data, but falls within conditional dist of y given x, we have a marginal y-outlier } not influential
- Influence depends on x-outlyingness (leverage) and conditional y-outlyingness (residual). Also sample size.
- Leverage values are bounded $\frac{1}{n} \le h_{ii} \le 1$
- Also, $\sum_{i=1}^n h_{ii} = p$, where $p = \text{rank}(X)$
  - ↳ so $\bar{h} = \frac{p}{n}$, and $2\bar{h}$ or $3\bar{h}$ are used as rules of thumb for high leverage.

Residuals:

- Using raw residuals for model diagnostics has limitations
- Raw res. do not all have same variance
- True errors: $Var(\varepsilon) = \sigma^2 I$; Res: $Var(e) = \sigma^2(I - H)$
  - ↳ So $Var(\varepsilon_i) = \sigma^2$, but $Var(e_i) = \sigma^2(1 - h_{ii})$
  - ↳ Var of res. for obs. depends on its leverage.
- We standardize them: $r_i = \frac{e_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}$ for constant variance 1.
  - ↳ but recall $\hat{\sigma} = \text{root MSE} = \sqrt{\frac{1}{n-p}\sum_{i=1}^n e_i^2}$ which depends on the raw residual
    1) If obs i is a regression outlier, and has an unusual res. it will inflate est of σ.
    2) Numerator and denominator not independent, so statistic doesn't follow a t dist.
- Studentized deleted residual: $d_i = \frac{e_i}{\hat{\sigma}_{-i}\sqrt{1 - h_{ii}}}$ where $\hat{\sigma}_{-i}$ is estimate of root MSE based on regression with observation i deleted.
  - ↳ If error terms normally dist, then stud. deleted res. have $t_{(n-1)-p}$ dist.
  - ↳ Don't refit model n times, just use relationship: $d_i = r_i\sqrt{\frac{n-p-1}{n-p-r_i^2}}$ (r_i standardized)

Component-plus-residuals plots

- AV plots not ideal for selecting transformations because it adjusts x_j for the other predictors
- Define partial residual for j^{th} predictor as $e_i^{(j)} = \hat{\beta}_j x_{ij} + e_i$
  - ↳ Add back linear component of partial relationship between Y and x_j to least-squares residual
- A component-plus-residual plot is one of $e_i^{(p-1)}$ versus $x_{i,p-1}$

Final Cubic Regression spline:

$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_{12} x_{i1}^2 + \beta_{13} x_{i1}^3 + \beta_{23} x_{i2}^3 + \beta_{33} x_{i3}^3$

where $x_{i1} \equiv x_i$

$x_{i2} = \begin{cases} 0 & \text{for } x_i \le c_1 \\ x_i - c_1 & \text{for } x_i > c_1 \end{cases}$

$x_{i3} = \begin{cases} 0 & \text{for } x_i \le c_2 \\ x_i - c_2 & \text{for } x_i > c_2 \end{cases}$

- If we fit a cubic regression spline w/ K knots, dividing data into $K+1$ bins, the model will have $K+4$ params.

$LOOCV_{(n)} = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_{i(-i)})^2 = \frac{1}{n}\sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{1 - h_i}\right)^2$

$CV_{(k\text{-fold})} = \frac{1}{k}\sum_{i=1}^k MSE_i$ → estimate for test MSE

↳ Natural or restricted cubic spline is constrained to be linear beyond left and right boundary knots

$MSE_{pred} = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_i)^2$

when ↓ training MSE, but ↑ test MSE, we are overfitting

heteroskedasticity seriously affects SEs only when ratio of largest to smallest variance is 4 or more

Measures of influence:

- Case deletion - how to calculate $\hat{\beta}$ with i^{th} case deleted?
  - ↳ Denote as $\hat{\beta}_{(-i)}$, a $p \times 1$ vector
  - ↳ X is $n \times p$ design matrix w/ linearly independent columns
  - ↳ $X_{(-i)}$ is design matrix with i^{th} obs deleted, a $(n-1) \times p$ matrix
  - ↳ $Y_{(-i)}$ is vector of observed values w/ i^{th} obs deleted; $(n-1) \times 1$
  - ↳ We want to compute $\hat{\beta}_{(-i)} = [X_{(-i)}'X_{(-i)}]^{-1}X_{(-i)}'Y_{(-i)}$
- Claim: we can compute $\hat{\beta}_{(-i)}$ for each i, to the original model output obtained by fitting the model to all n obs. without refitting the model
- $[X_{(-i)}'X_{(-i)}]^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}x_i x_i'(X'X)^{-1}}{1 - h_{ii}}$ where $h_{ii} = x_i'(X'X)^{-1}x_i$
- $X_{(-i)}'Y_{(-i)} = X'Y - x_i Y_i$
- $\hat{\beta}_{(-i)} = \left[(X'X)^{-1} + \frac{(X'X)^{-1}x_i x_i'(X'X)^{-1}}{1 - h_{ii}}\right][X'Y - x_i Y_i]$ $= \hat{\beta} - \frac{(X'X)^{-1}x_i e_i}{1 - h_i}$ → raw residual for obs. i
  - ↳ can obtain case-deleted MSE, fitted values, etc. using $\hat{\beta}_{(-i)}$. Don't refit the model every time.
- DFFITS: measures influence of i^{th} obs. on i^{th} fitted value $\hat{y}_i$
  - ↳ $DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i(-i)}}{\sqrt{\sigma^2_{(-i)}h_{ii}}}$ → fitted val for obs. i when i^{th} obs is omitted when fitting regression model
  - ↳ SD of $\hat{y}_i$, but using estimate of MSE when i^{th} obs. is omitted when fitting the model;

$Var(\hat{Y}_i) = Var(x_i'\hat{\beta})$
$= x_i' Var(\hat{\beta}) x_i$
$= \sigma^2 x_i'(X'X)^{-1}x_i$
$= \sigma^2 h_{ii}$

Also, $Var(\hat{Y}) = Var(HY) = H Var(Y) H = \sigma^2 H$

- Gives influence on fitted value in SD units
- Measure how well the model predicts each obs, when it is not used to fit the model. Think of them as measures of prediction accuracy.
- Note we can calculate using entire dataset:

$DFFITS_i = t_i\left(\frac{h_i}{1 - h_i}\right)^{1/2}$ where t_i is stud. deleted res. for obs. i and h_i is leverage value.

- Cook's Distance: summary index of influence of the i^{th} obs. on all fitted values:

$D_i = \frac{\sum_{j=1}^n (\hat{Y}_j - \hat{Y}_{j(-i)})^2}{p\hat{\sigma}^2} = \frac{(\hat{Y} - \hat{Y}_{(-i)})'(\hat{Y} - \hat{Y}_{(-i)})}{p\hat{\sigma}^2}$

$= \frac{(\hat{\beta}_{(-i)} - \hat{\beta})'X'X(\hat{\beta}_{(-i)} - \hat{\beta})}{p\sigma^2}$

$= \frac{e_i^2 h_i}{p\hat{\sigma}^2(1 - h_i)^2}$ (raw res.)

- DFBETAS: provide changes in j^{th} estimated regression coef when i^{th} obs. omitted from data set

$DFBETA_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{j(-i)}}{\sqrt{\sigma^2_{(-i)}C_{jj}}}$

where C_{jj} is j^{th} diagonal element of $(X'X)^{-1}$. Units in standard errors.

Regression model approach to RTM:

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
$= (\bar{Y} - \hat{\beta}_1\bar{x}) + \hat{\beta}_1 x_i$
$= \bar{Y} + \hat{\beta}_1(x_i - \bar{x})$ (centered predictor model)

↳ Now $\hat{\beta}_1 = r_{xy}\frac{s_Y}{s_X}$ so

$\hat{Y}_i = \bar{Y} + r_{xy}\frac{s_Y}{s_X}(x_i - \bar{x})$

$\frac{\hat{Y}_i - \bar{Y}}{s_Y} = r_{xy}\frac{x_i - \bar{x}}{s_x}$

- Because $0 < r_{xy} < 1$, predicted standardized value of Y is closer to mean Y than standardized value of x is to mean X. So $\hat{Y}_i$ will tend to be less extreme than x_i.

RTM occurs for any pair of variables. Happens when less than perfect correlation between variables.

- Percent regression to mean $= 100(1 - r)$

Addressing influential obs:

- Check for incorrect data
- Perhaps observations don't belong to pop. of interest
- May need to change model
- Consider impact of sample size.
- Sensitivity analysis → estimate model with and without point

- Bivariate p-value thresh
  - → arbitrary, doesn't deal w/ collinearity
- R^2 higher with more predictors
- R^2_{adj} larger for larger model whenever $F \ge 1$
  - ↳ selects too many variables
- Mallows C_p: Suppose we knew true model. measure diff between fitted vals from reduced $\hat{Y}_{iR}$ and true values Z_i as $K = \frac{1}{\sigma^2}\sum(\hat{Y}_{iR} - Z_i)^2$
  - K ↓ indicates good model $C_p = \frac{SSE(\ldots)}{\ldots} - (n - 2p)$

$AIC = -2\log_e L(\hat{\theta}) + 2p$ (likelihood evaluated at mle; → penalty for complexity)

↓ AIC is better

$= n\log_e(\hat{\sigma}^2) + 2p$

$\hat{\sigma}^2 = \frac{SSE}{n-p}$

$BIC = -2\log_e L(\hat{\theta}) + \log_e(n)p$

$= n\log_e(\hat{\sigma}^2) + \log_e(n)p$

AIC: ≤2 3-7 >10

SLR

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ → note, centered model gives $\hat{Y}_i = \bar{Y} + \hat{\beta}_1(X_i - \bar{X})$ ↳ because intercept becomes $\bar{Y}$

- β_0 = intercept; when range of data includes $X=0$, β_0 is the mean of the dist. of Y at $X=0$. Otherwise, not meaningful
- β_1 = slope of regression line; indicates change in conditional mean of Y per one unit increase in X
- σ^2 = conditional variance of Y given X; estimated by MSE

- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$
- $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$\hat{var}(\hat{\beta}_0) = \hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X_i - \bar{X})^2}\right) \Rightarrow \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{var}(\hat{\beta}_0)}} \sim t_{n-2}$; CI: $\hat{\beta}_0 \pm t_{1-\frac{\alpha}{2}, n-2}\, SE(\hat{\beta}_0)$

$\hat{var}(\hat{Y}_h) = \hat{\sigma}^2\left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right)$; CI: $\hat{Y}_h \pm t_{1-\frac{\alpha}{2}, n-2}\, SE(\hat{Y})$

prediction interval: $\hat{Y}_h \pm t_{1-\frac{\alpha}{2}, n-2}\, \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}}$ (different because we have two variance components: σ^2 and $\hat{var}(\hat{Y}_h)$)

- $\hat{\beta}_1 = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} = r\frac{S_Y}{S_X}$

$\hat{var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\Sigma(X_i - \bar{X})^2} \Rightarrow \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{var}(\hat{\beta}_1)}} \sim t_{n-2}$; CI: $\hat{\beta}_1 \pm t_{1-\frac{\alpha}{2}, n-2}\, SE(\hat{\beta}_1)$

- $\hat{\sigma}^2 = \frac{1}{n-2}\Sigma(Y_i - \hat{Y}_i)^2$; SSE (sum of squared errors) (residual df)
  - ↳ Mean square error, (MSE) error mean square, residual mean sum of squares (RSS)
- $\hat{\sigma}$ has same units as Y and is estimated std. dev. of observations around their mean.
  - ↳ root MSE
- Precision = $\frac{1}{\text{variance}}$

ANOVA: Explaining variation in Y

Total SS (TSS) $= \Sigma(Y_i - \bar{Y})^2$

$= \Sigma[(Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})]^2$

$= \underbrace{\Sigma(Y_i - \hat{Y}_i)^2}_{\text{Residual SS (RSS)}} + \underbrace{\Sigma(\hat{Y}_i - \bar{Y})^2}_{\text{Regression SS (RegSS)}}$

TSS = RSS + RegSS

$TSS = (n-1)S_Y^2$

- $R^2 = \frac{RegSS}{TSS} = 1 - \frac{RSS}{TSS}$ (coefficient of determination) (by linear rel. w/ regressors)
  - ↳ $0 \le R^2 \le 1$
  - ↳ % variation explained by model
  - ↳ may not capture strong non-linear relationship
  - ↳ high R^2 not a guarantee for good fit
- $r = \frac{\hat{cov}(x,y)}{\sqrt{\hat{var}(x)}\sqrt{\hat{var}(y)}} = \frac{\Sigma(x_i - \bar{x})(Y_i - \bar{Y})}{\sqrt{\Sigma(x_i - \bar{x})^2}\sqrt{\Sigma(Y_i - \bar{Y})^2}}$
  - ↳ for SLR, $r^2 = \pm R^2$
  - ↳ so for SLR, slope coefficient is a scaled version of r, by std. devs of x and y: $\hat{\beta}_1 = r\frac{S_Y}{S_X}$
  - ↳ $r_{x,y}$ values all given in Pearson correlation matrix for MLR (tells us about individual corr. matrix)

Source	df	SS	MS	F
Model	1	RegSS	MSReg = RegSS/1	MSReg / MSE
Residual	n-2	RSS	MSE = RSS/(n-2)	
Total	n-1	TSS		

$\sim F_{1, n-2}$

(note df are additive)

Note:
$E(MSE) = \sigma^2$
$E(MSReg) = \sigma^2 + \beta_1^2\Sigma(X_i - \bar{X})^2$
so as $\beta_1 \to 0$, MSReg → MSE

F-test gives same result as t-test for $\beta_1 = 0$
↳ large F therefore supports $\beta_1 \neq 0$.

From matrix: $Var(\hat{Y}) = Var(\hat{\beta}_0 + \hat{\beta}_1 10) = Var(\hat{\beta}_0) + 10^2 Var(\hat{\beta}_1) + 2(10)Cov(\hat{\beta}_0, \hat{\beta}_1)$ → For CI of point estimate of $\hat{Y}_h$

Assumptions:

- Linearity: $E(\varepsilon_i) \equiv E(\varepsilon_i | X_i) = 0$
  $E(Y_i) \equiv E(Y_i | X_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i) = \beta_0 + \beta_1 X_i + E(\varepsilon_i) = \beta_0 + \beta_1 X_i$
- Constant variance: $Var(\varepsilon | X_i) = \sigma^2$
  $Var(Y | X_i) = E[(Y_i - \mu_i)^2] = E[(Y_i - \beta_0 - \beta_1 X_i)^2] = E(\varepsilon_i^2) = \sigma^2$
- Normality: $\varepsilon_i \sim N(0, \sigma^2)$
  $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$
- Independence: Observations sampled independently. Must be evaluated by data collection procedure evaluation.
- Fixed X, or X measured without error and independent of errors.
- X is not invariant → must have variation in X pop.

MLR

For k predictors, model becomes:

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{ik} + \varepsilon_i$

↳ Note k+1 parameters

↳ β_j is change in $E(Y)$ for a one-unit increase in X_j when all other predictors are held constant. "controlling for all other predictors"

$R^2_{adj} = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)}$

- Note, models w/ higher R^2 may appear to be better fit, when it's really just an artifact of R^2 necessarily increasing w/ more predictors. So use adjusted R^2_{adj}

K = # predictors
k+1 = # parameters (β_0 term)

Degrees of freedom:
- RegSS has $(k+1) - 1 = k$ df
- RSS has $n - (k+1) = n - k - 1$ df
- TSS has $n - 1$ df

Omnibus: $H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$ (* All β_j's other than intercept is zero under H_0)
H_1: Not that

Source	df	SS	MS	F
Model	k	RegSS	RegSS/k	RegMS/MSE
Residual	n-k-1	RSS	RSS/(n-k-1)	
Total	n-1	TSS		

$\sim F_{k, n-k-1}$

↳ $F = \frac{RegSS/k}{RSS/(n-k-1)} = \left(\frac{n-k-1}{k}\right)\left(\frac{R^2}{1-R^2}\right)$

Test value of single coefficient:

$H_0: \beta_j = \beta_j^*$
$H_a: \beta_j \neq \beta_j^*$
$t^* = \frac{\hat{\beta}_j - \beta_j^*}{\sqrt{\hat{var}(\hat{\beta}_j)}} \sim t_{n-(k+1)}$

$\hat{var}(\hat{\beta}_j) = \left(\frac{1}{1-R_j^2}\right)\left(\frac{\sigma^2}{\Sigma_i(X_{ij} - \hat{X}_j)}\right)$

(R_j^2 from regression of X_j on all other Xs; σ^2 from MSE)

CI: $\hat{\beta}_j \pm t_{1-\frac{\alpha}{2}, n-k-1} * (SE(\hat{\beta}_j))$

Partial regression:
1. Regress Y on $X_2, \ldots, X_k$ and get residuals $e_{i, y|x_2, \ldots x_k}$
2. Regress X_1 on $X_2, \ldots, X_k$ and get residuals $e_{i, x_1|x_2, \ldots, x_k}$
3. SLR of $e_{i, y|x_2, \ldots, x_k}$ on $e_{i, x_1|x_2, \ldots, x_k}$ gives $\hat{\beta}_1$

* We think of each partial regression coefficient as providing an estimate of the linear relationship between $\hat{Y}$ and that predictor, after linear relationship between Y and all other predictors has been accounted (controlled) for.

Partial F-test

F^*

NOTE: df each equal to # model parameters limited in model

- If we want to test whether adding a set of predictors $X_m, X_{m+1}, \ldots, X_j$ to a model that already contains $X_1, X_2, \ldots, X_{m-1}$ has a sig. effect in terms of explaining variation in Y, we consider diff. in RSS or RegSS between models with and without $X_m, \ldots, X_j$:

  $SSE(X_1, \ldots, X_{m-1}) - SSE(X_1, \ldots, X_m, \ldots, X_j) = RegSS(X_1, \ldots, X_j) - RegSS(X_1, \ldots, X_{m-1})$

  * We are comparing a Full and a Reduced model! H_0: βs of removed vars = 0

Box-Cox transformations:

$$x \to x^{(p)} = \frac{x^p - 1}{p}$$

$x^{(0)}$ is $\ln x$ because $\lim_{p \to 0} \frac{x^p - 1}{p} = \ln x$

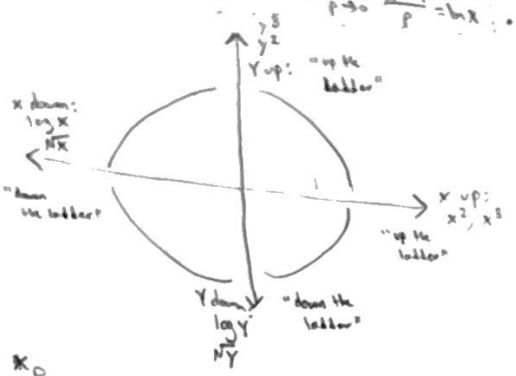
Transform X or Y?

- transforming Y generally has larger effect on dist. of error terms.
- think about interpretation

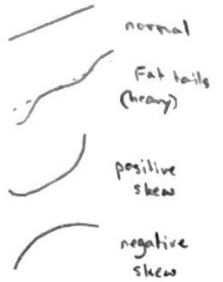
Model Diagnostics:

- Goal is to determine whether assumptions are met.
- Many assumptions are about the errors, which are not observable so we use residuals analysis instead.

Sometimes we look at standardized residuals: $e_i^* = \frac{e_i}{\text{MSE}}$ to give std. dev. = 1



Q-Q Interp



Q-Q has range of data

P-P has $[0,1] \times [0,1]$

Assumption	Methods for Checking	Consequence of violation	Remedies
constant variance	Residuals analysis, plot of e vs $\hat{Y}$	increased standard errors	variance stabilizing transformation of Y, weighted least squares, nothing if not bad
$E(\varepsilon_i) = 0$	Residuals analysis, plot of e vs. $\hat{Y}$	mean function is misspecified, regression coefficients are wrong	Transformations, deal w/ nonlinearity (i.e. polynomial reg.)
Normality (worst violation)	univariate residuals plot; Q-Q, P-P	important for validity of tests and CI in small samples	Transformation of Y
Independent obs.	understand data collection	SEs are wrong (↓) if we model correlated values as though independent; parameter estimates biased	methods for correlated data

* Power transformations can help linearize when relationship is monotone and "simple" (i.e. no point of inflection). And all values are positive → add a start if need be

In general, for $\ln(y)$, effect of a one-unit increase in X_j is to multiply mean of Y by $e^{\hat{\beta}_j}$

Interpreting natural log transformations: When we ln-transform a predictor, we can say that a 1% increase in x is associated with an increase in the conditional mean of y by about $0.01\hat{\beta}$. When we ln-transform the outcome variable, the effect of a one-unit increase in x is to multiply y by $\exp[\hat{\beta}_1]$.

Suppose y and x are related by $y = \alpha x^{\beta}$. If we increase x by 1%, the change in y is equivalent to multiplying it by 1.01^{β}, which is roughly equal to increasing y by β percent. ⇝ When y and x are both ln-transformed. β is here called an elasticity coefficient

Distinction between correlation and interaction:

- Moderation concerns how two or more predictor variables together affect a third variable, the outcome
- Interaction and correlation are often confused, but:
  - X_1 and X_2 interact to affect Y
  - X_1 and X_2 are correlated

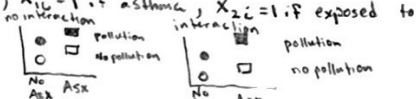
- Correlation between two variables measures the degree of linear association between them
- Interaction between two variables means that the effect of one of the variables on a third variable depends on the value of the second variable.

* Variables can be correlated but not interact and vice versa

Suppose Y_i is measure of breathing difficulty, $X_{1i} = 1$ if asthma, $X_{2i} = 1$ if exposed to pollution

Interaction model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \varepsilon_i$$



↳ For children w/ asthma, mean change due to air pollution exposure is $\beta_2 + \beta_3$

↳ For no asthma, mean change due to air pollution exposure is β_2

*We say "asthma modifies (moderates) the effect of air pollution on breathing difficulties" or vice versa.

OR "Effect of air pollution on breathing difficulties is conditional on asx status"

Interaction terms are symmetric in the two variables:

↳ $Y_i = \beta_0 + \beta_2 X_{2i} + (\beta_1 + \beta_3 X_{2i}) X_{1i} + \varepsilon_i$

- When $X_{2i} = 0$, effect of X_{1i} on $E(Y_i)$ is β_1
- When $X_{2i} = 1$, effect of X_{2i} on $E(Y_i)$ is $\beta_1 + \beta_3$

↳ $Y_i = \beta_0 + \beta_1 X_{1i} + (\beta_2 + \beta_3 X_{1i}) X_{2i} + \varepsilon_i$

- When $X_{1i} = 0$, effect of X_{2i} on $E(Y_i)$ is β_2
- When $X_{1i} = 1$, effect of X_{2i} on $E(Y_i)$ is $\beta_2 + \beta_3$

We can have interactions among 3 or more variables.

Here's a model with 3 dichotomous (0/1) variables and a 3-way interaction:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_3 + \beta_4 X_{1i} X_{2i} + \beta_5 X_{1i} X_{3i}$$

If one predictor was categorical w/ 3 levels? Let Z_1 and Z_2 be dummy variables:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_{1i} + \beta_3 Z_{2i} + \beta_4 X_i Z_{1i} + \beta_5 X_i Z_{2i} + \varepsilon_i$$

When we have more than one interaction term, we conduct a partial F-test as to whether or not an interaction exists: $H_0: \beta_4 = 0, \beta_5 = 0$

Testing:

- Higher order terms first, as in polynomial regression
- Do not drop lower-order terms that are "not significant"
- Note nesting of diff slope and intercept model and common slope model. i.e. test $\beta_3 = 0$

Interaction between categorical and continuous variable

No interaction: $d_i = 1$ for men $d_i = 0$ for women; X_i yrs ed. Y_i income in $

$Y_i = \beta_0 + \beta_1 d_i + \beta_2 X_i + \varepsilon_i$

↳ women: $\beta_0 + \beta_2 X_i + \varepsilon_i$

↳ men: $(\beta_0 + \beta_1) + \beta_2 X_i + \varepsilon_i$

* common slope model

↳ β_1 is Δ intercepts, constant vertical dist. btwn lines, expected difference in mean income btwn men and women when education held constant.

Interaction:

$Y_i = \beta_0 + \beta_1 d_i + \beta_2 X_i + \beta_3 d_i X_i + \varepsilon_i$

↳ women: $\beta_0 + \beta_2 X_i + \varepsilon_i$

↳ men: $(\beta_0 + \beta_1) + (\beta_2 + \beta_3) X_i + \varepsilon_i$

* Allows for difference in slope!

β_0 is int. for women

β_1 is Δ int. for men vs. women

β_2 is slope for women

β_3 Δ slope for men vs. women

Interpretation of lower-order terms in model w/ interaction term

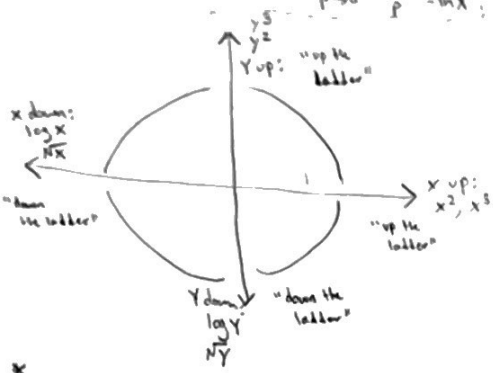
$Y_i = \beta_0 + \beta_1 d_i + (\beta_2 + \beta_3 d_i) X_i + \varepsilon_i$

- β_1 is diff. in mean outcome for individuals who differ on d_i by 1 unit and have $X_i = 0$ → if this doesn't make sense, center X_i at its mean!

Box-Cox transformation:

$$x \to x^{(p)} = \frac{x^p - 1}{p}$$

$x^{(0)}$ is $\ln x$ because $\lim_{p \to 0} \frac{x^p - 1}{p} = \ln x$



*Power transformations can help linearize when relationship is monotone and "simple" (i.e. no point of inflection). And all values are positive → add a start if need be

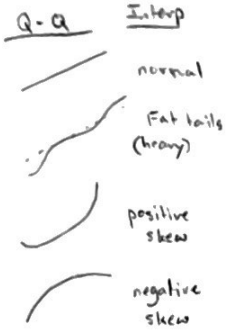
Transform X or Y?

- transforming Y generally has larger effect on dist. of error terms.
- think about interpretation

Model Diagnostics:

- Goal is to determine whether assumptions are met.
- Many assumptions are about the errors, which are not observable so we use residuals analysis instead.

Sometimes we look at standardized residuals: $e_i^* = \frac{e_i}{s\sqrt{1-h_{ii}}}$ to give std. dev. of 1



Q-Q has range of data
P-P has $[0,1] \times [0,1]$

Assumption	Methods for Checking	Consequence of violation	Remedies
constant variance	Residuals analysis, plot of e vs $\hat{Y}$	increased standard errors	variance stabilizing transformation of Y, weighted least squares, nothing if not bad
$E(\varepsilon_i) = 0$	Residuals analysis, plot of e vs. $\hat{Y}$	mean function is misspecified, regression coefficients are wrong	Transformations, deal w/ non-linear (i.e. polynomial reg.)
Normality (worst violation)	univariate residuals plot; Q-Q, P-P	important for validity of tests and CI in small samples	Transformation of Y
Independent obs.	Understand data collection	SEs are wrong (↓) if we model correlated values as though independent; parameter estimates biased	methods for correlated data

In general, for $\ln(y)$, effect of a one-unit increase in X_j is to multiply mean of Y by $e^{\hat{\beta}_j}$

Interpreting natural log transformations: When we ln-transform a predictor, we can say that a 1% increase in x is associated with an increase in the conditional mean of y by about $0.01\hat{\beta}$. When we ln-transform the outcome variable, the effect of a one-unit increase in x is to multiply y by $\exp[\hat{\beta}_1]$.

Suppose y and x are related by $y = \alpha x^\beta$. If we increase x by 1%, the change in y is equivalent to multiplying it by 1.01^β, which is roughly equal to increasing y by β percent. ↝ When y and x are both ln-transformed, β is here called an elasticity coefficient

Distinction between correlation and interaction:

- Moderation concerns how two or more predictor variables together affect a third variable, the outcome
- Interaction and correlation are often confused, but:
  - X_1 and X_2 interact to affect Y
  - X_1 and X_2 are correlated

- Correlation between two variables measures the degree of linear association between them
- Interaction between two variables means that the effect of one of the variables on a third variable depends on the value of the second variable

* Variables can be correlated but not interact and vice versa

Testing:

- Higher order terms first, as in polynomial regression
- Do not drop lower-order terms that are "not significant"
- Note nesting of diff slope and intercept model and common slope model. i.e. test $\beta_3 = 0$

Interaction between categorical and continuous variable

No interaction: $d_i = 1$ for men $d_i = 0$ for women; x_i yrs ed. Y_i income in $

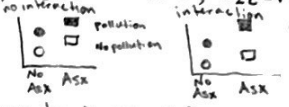
$Y_i = \beta_0 + \beta_1 d_i + \beta_2 x_i + \varepsilon_i$

↳ women: $\beta_0 + \beta_2 x_i + \varepsilon_i$
↳ men: $(\beta_0 + \beta_1) + \beta_2 x_i + \varepsilon_i$ * common slope model

↳ β_1 is Δ intercepts, constant vertical dist. btwn lines, expected difference in mean income btwn men and women when education held constants.

Interaction:
$Y_i = \beta_0 + \beta_1 d_i$ for is int. for women
men vs. women
vs. women

Suppose Y_i is measure of breathing difficulty, $X_{1i} = 1$ if asthma, $X_{2i} = 1$ if exposed to pollution
Interaction model:



$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \varepsilon_i$$

↳ For children w/ asthma, mean change due to air pollution exposure is $\beta_2 + \beta_3$
↳ For no asthma, mean change due to air pollution exposure is β_2

*We say "asthma modifies (moderates) the effect of air pollution on breathing difficulties" or vice versa.

OR "Effect of air pollution on breathing difficulties is conditional on asx status"

Interaction terms are symmetric in the two variables:

↳ $Y_i = \beta_0 + \beta_2 X_{2i} + (\beta_1 + \beta_3 X_{2i}) X_{1i} + \varepsilon_i$
- When $X_{2i} = 0$, effect of X_{1i} on $E(Y_i)$ is β_1
- When $X_{2i} = 1$, effect of X_{2i} on $E(Y_i)$ is $\beta_1 + \beta_3$

↳ $Y_i = \beta_0 + \beta_1 X_{1i} + (\beta_2 + \beta_3 X_{1i}) X_{2i} + \varepsilon_i$
- When $X_{1i} = 0$, effect of X_{2i} on $E(Y_i)$ is β_2
- When $X_{1i} = 1$, effect of X_{2i} on $E(Y_i)$ is $\beta_2 + \beta_3$

We can have interactions among 3 or more variables.
Here's a model with 3 dichotomous (0,1) variables and a 3-way interaction:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_3 + \beta_4 X_{1i} X_{2i} + \beta_5 X_{1i} X_{3i}$$

If one predictor was categorical w/ 3 levels? Let Z_1 and Z_2 be dummy variables:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_{1i} + \beta_3 Z_{2i} + \beta_4 X_i Z_{1i} + \beta_5 X_i Z_{2i} + \varepsilon_i$$

When we have more than one interaction term, we conduct a partial F-test as to whether or not an interaction exists: $H_0: \beta_4 = 0, \beta_5 = 0$

Interpretation of lower-order terms in model w/ interaction term
$Y_i = \beta_0 + \beta_1 d_i + (\beta_2 + \beta_3 d_i) X_i + \varepsilon_i$
- β_1 is diff. in mean outcome for individuals who differ on d_i by 1 unit and have $X_i = 0$ → if this doesn't make sense, center X_i at its mean!

Inequalities

A function f(x, y) is a joint pdf ... for every $A \subset \mathbb{R}^2$, $P((X,Y)\in A) = \int_A \int f(x,y)\,dx\,dy$

$X \perp Y$ if for all x, y $f(x,y) = f(x)f(y)$

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

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

Conditionals:

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

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

→ This implies that the conditionals equal the marginals when $X \perp Y$

• Suppose mgf $M(t_1,t_2)$ exists for r.v.s X_1 and X_2. $X_1 \perp X_2$ iff $M(t_1,t_2) = M(t_1,0)M(0,t_2)$

• Cauchy-Schwarz

For any 2 r.v.s X and Y

$|EXY| \le E|XY| \le (E|X|^2)^{1/2}(E|Y|^2)^{1/2}$

So, if X and Y have means μ_X and μ_Y and variances σ_X^2 and σ_Y^2, we have

$(Cov(X,Y))^2 \le \sigma_X^2 \sigma_Y^2 \Rightarrow$ proves $\rho^2 < 1$

• Jensen's (if f is differentiable, f convex iff $f'(x) \le f'(y)$ for all $a<x<y<b$; if twice differentiable, f convex iff $f''(x) \ge 0$ for all $a<x<y<b$)

For any r.v. X, if $g(x)$ is convex, then $Eg(X) \ge g(EX)$

; if $g(x)$ is concave, then $Eg(X) \le g(EX)$

Equality holds if and only if, for every line $a+bx$ tangent to $g(x)$ at $x = EX$, $P(g(X) = a+bX) = 1$

* One immediate application tells us $EX^2 \ge (EX)^2$, since $g(x) = x^2$ is convex

** Strict inequality holds if $g(x)$ is strictly convex

• Chebyshev's

Let X be a r.v. and let $g(x)$ be a nonnegative. Then for any $r>0$,

$P(g(X) \ge r) \le \frac{Eg(X)}{r}$

* Most common use says to let $g(x) = \frac{(x-\mu)^2}{\sigma^2}$, $r = t^2$. Then $P\left(\frac{(X-\mu)^2}{\sigma^2} \ge t^2\right) \le \frac{1}{t^2}E\left(\frac{(X-\mu)^2}{\sigma^2}\right) = \frac{1}{t^2}$. For convenience, rewrite

• Markov's

Let $u(x)$ be a nonnegative function of the r.v. X. If $E[u(X)]$ exists, then for every positive constant c, $P[u(X) \ge c] \le \frac{E[u(X)]}{c}$

• $Eg(X) = \begin{cases} \int_{-\infty}^{\infty} g(x)f(x)\,dx, & \text{if X is continuous} \\ \sum_{x\in\mathcal{X}} g(x)p(x), & \text{if X is discrete} \end{cases}$

• For any constants a and b, the mgf of the r.v. $aX+b$ is given by $M_{aX+b}(t) = e^{bt}M_X(at)$

• $E[ag_1(X) + bg_2(X) + c] = aEg_1(X) + bEg_2(X) + c$

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

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

• If $X \perp Y$, $E[g(X)h(Y)] = [Eg(X)][Eh(Y)]$

• $Var(Y|X) = E(Y^2|X) - E(Y|X)^2$

• If $X \perp Y$, and $Z = X+Y$, then $M_Z(t) = M_X(t)M_Y(t)$

• If X and Y are any two r.v.s, $EX = E[E(X|Y)]$

• For any two r.v.s X and Y, $Var X = E[Var(X|Y)] + Var[E(X|Y)]$

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

• $\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$ * If $X \perp Y$, then $Cov(X,Y) = \rho = 0$

• $Var(aX+bY) = a^2 Var X + b^2 Var Y + 2ab\,Cov(X,Y)$ * So when $X \perp Y$, $Var(aX+bY) = a^2 Var X + b^2 Var Y$

General Mathematical Facts:

$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$, $|x|<1$

$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ $\frac{1}{1+x} = \sum_{n=0}^{\infty} (-1)^n x^n$

$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$

$(fg)' = f'g + fg'$

$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$ $\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1}e^{-y}\,dy$

$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} b^x a^{n-x}$

$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n+1}\frac{x^n}{n}$

$(f(g(x)))' = f'(g(x))g'(x)$

$(f(x) \pm g(x))' = f'(x) \pm g'(x)$

$\frac{a}{1-r} = \sum_{n=1}^{\infty} ar^{n-1}$ for $|r|<1$, otherwise diverges

$\frac{d}{dx}(a^x) = a^x \ln(a)$ $\frac{d}{dx}(\ln x) = \frac{1}{x}, x>0$

$\frac{d}{dx}(e^x) = e^x$ $\frac{d}{dx}(\ln|x|) = \frac{1}{x}, x \ne 0$

Maclaurin series for mgf $M(t)$:

$M(t) = M(0) + \frac{M'(0)}{1!}t + \frac{M''(0)}{2!}t^2 + \dots + \frac{M^{(m)}(0)}{m!}t^m$

$\frac{d}{dx}(\log_a x) = \frac{1}{x\ln a}$

$\int x^{-n}\,dn = \left(\frac{1}{1-n}\right)x^{-n+1}$

$\int e^u\,du = e^u$; $\int a^u\,du = \frac{a^u}{\ln a}$ $n \ne 1$; $\int \frac{1}{ax+b}\,dx = \frac{1}{a}\ln|ax+b|$; $\int x^{p/q}\,dx = \frac{q}{p+q}x^{\frac{p+q}{q}}$ for $x>0$

$\int \frac{1}{u\ln u}\,du = \ln|\ln u|$; $\int \ln u\,du = u\ln u - u$; $\int ue^u\,du = (u-1)e^u$

Discrete Uniform

pmf: $p(x) = \frac{1}{N}$; $x = 1,2,\dots$; $N = 1,2,\dots$

mgf: $M(t) = \left(\frac{1}{N}\right)\sum_{i=1}^{N} e^{it}$

$\mu = \frac{N+1}{2}$; $\sigma^2 = \frac{(N+1)(N-1)}{12}$

Bernoulli: bernoulli(p)

pmf: $p(x) = p^x(1-p)^{1-x}$; $x = 0,1$; $0 \le p \le 1$

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

$\mu = p$; $\sigma^2 = p(1-p)$

Binomial binomial(n, p)

\# successes in n independent Bernoulli trials.

pmf: $p(x) = \binom{n}{x}p^x(1-p)^{n-x}$, $x = 0,1,\dots,n$

mgf: $M(t) = E(e^{tX}) = \sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x}e^{tx}$

$= \sum_{x=0}^{n}\binom{n}{x}(e^tp)^x(1-p)^{n-x}$

$= [(e^tp) + (1-p)]^n$

$\mu = np$; $\sigma^2 = np(1-p)$

** $X_1, X_2, \dots, X_m$ are independent r.v.s such that $X_i \sim b(n_i, p)$. Let $Y = \sum_{i=1}^{m} X_i$, then $Y \sim b\left(\sum_{i=1}^{m} n_i, p\right)$

Multinomial

Experiment results in one of k ways, $C_1, \dots, C_k$. p_i is prob. of result being an element of C_i. Experiment repeated n times. X_i is number of outcomes in C_i. If $X_1, X_2, \dots, X_{k-1}$ are # of outcomes in corresponding set C_i, then the multinomial pmf is given by:

$p(x) = \frac{n!}{x_1!\cdots x_{k-1}!\,x_k!}p_1^{x_1}p_2^{x_2}\cdots p^{x_{k-1}}p^{x_k}$

Specifically, when $k=3$, we get the trinomial

$p(x,y) = \frac{n!}{x!\,y!\,(n-x-y)!}p_1^x p_2^y p_3^{n-x-y}$

$M(t_1,t_2) = (p_1e^{t_1} + p_2e^{t_2} + p_3)^n$

$M(t_1,0) = M_X(t_1) = [(1-p_1) + p_1e^{t_1}]^n$

$M(0,t_2) = M_Y(t_2) = [(1-p_2) + p_2e^{t_2}]^n$

Negative Binomial (r, p)

\# failures in a sequence of independent Bernoulli trials before the occurrence of the rth success.

pmf: $p(x) = \binom{x+r-1}{r-1}p^r(1-p)^x$

OR $p(y) = \binom{y-1}{r-1}p^r(1-p)^{y-r}$ where $x = 0,1,\dots$, $y = r, r+1, \dots$

mgf: $M(t) = \left[\frac{p}{1-(1-p)e^t}\right]^r$, $t < -\log(1-p)$

$\mu = \frac{r(1-p)}{p}$; $\sigma^2 = \frac{r(1-p)}{p^2}$

* Can be derived as a gamma mixture of Poissons

Geometric

A special case of the negative binomial distribution with $r=1$. geometric(p)

$p(x) = p(1-p)^x$, $x = 0,1,\dots$ $\mu = \frac{1}{p}$

mgf: $M(t) = \frac{pe^t}{1-(1-p)e^t}$, $t < -\log(1-p)$ $\sigma^2 = \frac{1-p}{p^2}$

* This distribution is memoryless: $P(X>s|X>t) = P(X>s-t)$

* $Y = X-1$ is negative binomial(1, p)

Hypergeometric

\# of defective items in a sample of size K, supposing we have a lot of N items in which M are defective

pmf: $p(x) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$, $M-(N-K) \le x \le M$, $N, M, K \ge 0$, $x = 0,1,\dots,k$

$\mu = \frac{KM}{N}$; $\sigma^2 = \left(\frac{KM}{N}\right)\left[\frac{(N-M)(N-K)}{N(N-1)}\right]$

Poisson(λ)

\# of events occurring in a fixed interval

pmf: $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0,1,\dots$; $0 \le \lambda < \infty$

mgf: $M(t) = e^{\lambda(e^t-1)}$; $\mu = \sigma^2 = \lambda$

Gamma(α, β)

Frequently a model for waiting times.

pdf: $f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}$, $0 \le x < \infty$, $\alpha, \beta > 0$

mgf: $M(t) = \left(\frac{1}{1-\beta t}\right)^\alpha$, $t < \frac{1}{\beta}$

* $\alpha = 1$ gives exponential $\mu = \alpha\beta$

* $\alpha = p/2$, $\beta = 2$ gives χ^2 $\sigma^2 = \alpha\beta^2$

Exponential(β)

Frequently a model for waiting time until first change.

pdf: $f(x) = \frac{1}{\beta}e^{-x/\beta}$, $0 \le x < \infty$, $\beta > 0$

mgf: $M(t) = \frac{1}{1-\beta t}$, $t < 1/\beta$

* special case of gamma $\mu = \beta$

* memoryless (like geometric) $\sigma^2 = \beta^2$

Lognormal(μ, σ²)

pdf: $f(x) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)\left(\frac{e^{-(\log x - \mu)^2/2\sigma^2}}{x}\right)$

$0 \le x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$

mgf dne. Moments: $EX^n = e^{n\mu + n^2\sigma^2/2}$

$\mu = e^{\mu + (\sigma^2/2)}$; $\sigma^2 = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$

Chi squared (p)

pdf: $f(x) = \frac{1}{\Gamma(p/2)2^{p/2}}x^{(p/2)-1}e^{-x/2}$

$0 \le x < \infty$; $p = 1,2,3,\dots$

mgf: $M(t) = \left(\frac{1}{1-2t}\right)^{p/2}$, $t < \frac{1}{2}$

$\mu = p$; $\sigma^2 = 2p$

* Special case of gamma

Beta(α, β)

pdf: $\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$, $0 \le x \le 1$, $\alpha > 0$, $\beta > 0$

mgf: $M(t) = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!}$ $\mu = \frac{\alpha}{\alpha+\beta}$

$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

* The constant in beta pdf can be defined in terms of gamma functions, $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

Normal(μ, σ²)

pdf: $f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$

$-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$

mgf: $M(t) = e^{\mu t + \sigma^2t^2/2}$

$\mu = \mu$, $\sigma^2 = \sigma^2$

Uniform(a, b)

pdf: $f(x) = \frac{1}{b-a}$, $a \le x \le b$

mgf: $M(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$

$\mu = \frac{b+a}{2}$

$\sigma^2 = \frac{(b-a)^2}{12}$

* If $a=0$, $b=1$, this is special case of beta

F

Related to $\chi^2 \Rightarrow F_{\nu_1,\nu_2} = \left(\frac{\chi^2_{\nu_1}}{\nu_1}\right)\Big/\left(\frac{\chi^2_{\nu_2}}{\nu_2}\right)$ where the χ^2 are independent. Also related to $t \Rightarrow F_{1,\nu} = t_\nu^2$

$\mu = \frac{\nu_2}{\nu_2-2}$, $\nu_2 > 2$; $\sigma^2 = 2\left(\frac{\nu_2}{\nu_2-2}\right)^2\frac{\nu_1+\nu_2-2}{\nu_1(\nu_2-4)}$, $\nu_2 > 4$

pdf: $f(x) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}\frac{x^{(\nu_1-2)/2}}{\left(1+\left(\frac{\nu_1}{\nu_2}\right)x\right)^{(\nu_1+\nu_2)/2}}$

$0 \le x < \infty$, $\nu_1, \nu_2 = 1,2,\dots$

mgf dne Moments: $EX^n = \frac{\Gamma\left(\frac{\nu_1+2n}{2}\right)\Gamma\left(\frac{\nu_2-2n}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_2}{\nu_1}\right)^n$, $n < \frac{\nu_2}{2}$

Cauchy(θ, σ)

$f(x) = \left(\frac{1}{\pi\sigma}\right)\left(\frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2}\right)$

$-\infty < x < \infty$; $-\infty < \theta < \infty$, $\sigma > 0$

mgf, μ, σ^2 dne

* special case of t, when df = 1. Also if X, Y iid n(0,1), then X/Y is cauchy

t

Related to $F \Rightarrow F_{1,\nu} = t_\nu^2$

$\mu = 0$, $\nu > 1$; $\sigma^2 = \frac{\nu}{\nu-2}$, $\nu > 2$

pdf: $f(x) = \left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)\left(\frac{1}{\sqrt{\nu\pi}}\right)\left(\frac{1}{\left(1+\left(\frac{x^2}{\nu}\right)\right)^{(\nu+1)/2}}\right)$

mgf dne

Moments:

$EX^n = 0$ if $n < \nu$ and odd

$EX^n = \left(\frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{\nu-n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\right)\nu^{n/2}$ if $n < \nu$ and even

Bivariate normal

$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-q/2}$

where $q = \left(\frac{1}{1-\rho^2}\right)\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$

•If $X \sim N(\mu, \sigma^2)$, then $V = (X-\mu)^2/\sigma^2 \sim \chi^2(1)$

•Let $X_1, \dots, X_n$ be ⊥ r.v.s so for $i = 1,2,\dots,n$, $X_i \sim N(\mu_i, \sigma_i^2)$. Let $Y = \sum_{i=1}^{n} a_iX_i$, for constants $a_1, \dots, a_n$.

$Y \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma^2\right)$

•Let $X_1, \dots, X_n$ be iid so $X_i \sim N(\mu, \sigma^2)$. Then $\bar{X} \sim N(\mu, \sigma^2/n)$

•Let $X \sim \chi^2(r)$. If $k > -r/2$, then $E(X^k)$ exists and is given by $E(X^k) = \frac{2^k\Gamma\left(\frac{r}{2}+k\right)}{\Gamma\left(\frac{r}{2}\right)}$ (if $k > -r/2$)

•Let $X_1, \dots, X_n$ be iid. Suppose $X_i \sim \Gamma(\alpha_i, \beta)$. Let $Y = \sum_{i=1}^{n} X_i$. Then $Y \sim \Gamma\left(\sum_{i=1}^{n}\alpha_i, \beta\right)$

•Let $X_1, \dots, X_n$ be ⊥ r.v.s. Suppose $X_i \sim \chi^2(r_i)$. Let $Y = \sum_{i=1}^{n} X_i$. Then $Y \sim \chi^2\left(\sum_{i=1}^{n} r_i\right)$

• We get a t-distribution when $W \sim N(0,1)$, $V \sim \chi^2(r)$ and we take $T = \frac{W}{\sqrt{V/r}}$

• We get an F-distribution when $U \sim \chi^2(r_1)$ and $V \sim \chi^2(r_2)$ and we take $W = \frac{(U/r_1)}{(V/r_2)}$

• $X_1, \dots, X_n$ are iid $\sim N(\mu, \sigma^2)$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(X_i - \bar{X})^2$

a) $\bar{X} \sim N(\mu, \sigma^2/n)$ b) $\bar{X} \perp S^2$

c) $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$

d) $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ d.f.

$\int\int f(x_1,x_2)\,dx_1\,dx_2 = \int\int g(y_1,y_2)|J|\,dy_1\,dy_2$ [illegible]

Bivariate trans. We start w/ (X_1, X_2) and want (Y_1, Y_2) in terms of X. We make sure ... express x_1 and x_2 in terms of y, and to transform Y_2. Get inverses, $w_1(y_1,y_2)$, $w_2(y_1,y_2)$. Then get Jacobian. ... old bounds ...

Commutativity:
$A \cup B = B \cup A$
$A \cap B = B \cap A$

Associativity:
$A \cup (B \cup C) = (A \cup B) \cup C$
$A \cap (B \cap C) = (A \cap B) \cap C$

Distributive laws:
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

DeMorgan's laws:
$(A \cup B)^c = A^c \cap B^c$
$(A \cap B)^c = A^c \cup B^c$

$P(\text{full house}) = \dfrac{\binom{13}{2}\binom{4}{0}\binom{4}{3}\binom{1}{0}\binom{4}{3}\binom{4}{2} \binom{1}{0}\binom{1}{0}}{\binom{52}{5}}$

	with replacement	without replacement
ordered	n^r	$\frac{n!}{(n-r)!}$
unordered	$\binom{n+r-1}{r}$	$\binom{n}{r} = \frac{n!}{r!(n-r)!}$

Basic properties of $P(\cdot)$:
① $P(\emptyset) = 0$
② $0 \le P(A)$
③ $P(A^c) = 1 - P(A)$
④ If $A_1 \subset A_2$, then $P(A_1) \le P(A_2)$
⑤ $P(B \cap A^c) = P(B) - P(A \cap B)$
⑥ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
↳ $P(B \cup A) \le P(A) + P(B)$ "Boole's"
↳ $P(A^c \cap B^c) \ge 1 - (P(A) + P(B))$

Conditional probability and independence:
Def'n: $P(A \cap B) = P(A)P(B)$ means statistical independence
↳ from Bayes' w/ assumption that given has no effect
↳ If A and B are independent, so are
- A and B^c
- A^c and B
- A^c and B^c

Def'n conditional prob: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$
↳ $P(A \cap B) = P(A|B)P(B)$
↳ $P(A \cap B) = P(B|A)P(A)$
↳ $P(A|B) = P(B|A)\left(\dfrac{P(A)}{P(B)}\right)$ "Bayes"

cdf of X defined by $F_X(x) = P(X \le x)$ for all x
$F_X(x)$ is a cdf iff
a) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$
b) $F_X(x)$ is a nondecreasing function of x
c) $F_X(x)$ is right-continuous: for every x_0, $\lim_{x \downarrow x_0} F_X(x) = F(x_0)$

The following two statements are equivalent:
1) r.v.s X and Y are identically distributed
2) $F_X(x) = F_Y(x)$ for every x

pmf of a discrete r.v. X given by $f_X(x) = P(X = x)$ for all x

pdf of a continuous r.v. X is the function that satisfies $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ for all x
$P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ for all x
Thus, $\frac{d}{dx}F_X(x) = f_X(x)$

$f_X(x)$ is a pdf or pmf iff:
a) $f_X(x) \ge 0$ for all x
b) $\sum_x f_X(x) = 1$ (pmf) OR $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ (pdf)

Expectation:
$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$ OR $E(X) = \sum_x x f(x)$
$E(k_1 g_1(x) + k_2 g_2(x) + c) = k_1 E(g_1(x)) + k_2 E(g_2(x)) + c$

LotUS: Suppose X is a continuous r.v. defined by pdf $f(x)$. Let Y be a r.v. $Y = g(X)$ for some function g.
$E(Y) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

Suppose (X_1, X_2) is a continuous random vector. Let Y be a r.v. defined by $Y = g(X_1, X_2)$ for some real-valued function g (i.e. $\mathbb{R}^2 \to \mathbb{R}$). Then $E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1,x_2) f(x_1,x_2)\,dx_1\,dx_2$ (↳ joint pdf)
$E(k_1 Y_1 + k_2 Y_2) = k_1 E(Y_1) + k_2 E(Y_2)$

Variance: Let X be a r.v. with finite mean $\mu = E(X)$ and such that $E((X-\mu)^2)$ is finite. Then the variance of X is: $\sigma^2 = E((X-\mu)^2)$
$= E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2$
$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$

Multivariate:
Joint cdf: $P(a_1 < X_1 \le b_1,\ a_2 < X_2 \le b_2) = F_{X_1,X_2}(b_1,b_2) - F_{X_1,X_2}(a_1,b_2) - F_{X_1,X_2}(b_1,a_2) + F_{X_1,X_2}(a_1,a_2)$

Joint pmf: $p_{X_1,X_2}(x_1,x_2) = P(X_1 = x_1, X_2 = x_2)$

Joint pdf: $F_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{X_1,X_2}(w_1,w_2)\,dw_1\,dw_2$
$\dfrac{\partial^2 F_{X_1,X_2}(x_1,x_2)}{\partial x_1 \partial x_2} = f_{X_1,X_2}(x_1,x_2)$

Note that $P((X_1,X_2) \in A) = \iint_A f_{X_1,X_2}(x_1,x_2)\,dx_1\,dx_2$, which is just the volume under the surface $z = f_{X_1,X_2}(x_1,x_2)$ over the set A.

Marginals: To find marginal pmfs, hold marginal variable and sum over the other.
To find marginal pdfs, $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1,x_2)\,dx_2$ (joint w/ respect to other r.v.)
To find marginal cdfs, $F_{X_1}(x_1) = P(X_1 \le x_1, -\infty < X_2 < \infty) = \lim_{x_2 \to \infty} F(x_1,x_2)$

Conditional distributions and expect.:
Discrete: $P(X_2 = x_2 | X_1 = x_1) = \dfrac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)}$
$= \dfrac{p_{X_1X_2}(x_1,x_2)}{p_{X_1}(x_1)} = p_{X_2|X_1}(x_2|x_1)$ ↗ conditional pmf

Continuous: $f_{X_2|X_1}(x_2|x_1) = \dfrac{f_{X_1X_2}(x_1,x_2)}{f_{X_1}(x_1)}$
Note: $\int_{-\infty}^{\infty} f_{X_2|X_1}(x_2|x_1)\,dx_2 = \int_{-\infty}^{\infty} \dfrac{f_{X_1X_2}(x_1,x_2)}{f_{X_1}(x_1)}\,dx_2$
$= \dfrac{1}{f_{X_1}(x_1)}\int_{-\infty}^{\infty} f_{X_1X_2}(x_1,x_2)\,dx_2 = \dfrac{f_{X_1}(x_1)}{f_{X_1}(x_1)} = 1$

$P(a < X_2 < b | X_1 = x_1) = \int_a^b f_{2|1}(x_2|x_1)\,dx_2$
$P(c < X_1 < d | X_2 = x_2) = \int_c^d f_{1|2}(x_1|x_2)\,dx_1$

If $u(X_2)$ is a function of X_2, the conditional expectation of $u(X_2)$, given that $X_1 = x_1$, is
$E(u(X_2)|x_1) = \int_{-\infty}^{\infty} u(x_2) f_{2|1}(x_2|x_1)\,dx_2$
Conditional variance: $E((X_2 - E(X_2|x_1))^2 | x_1)$
Conditional mean: $E(X_2|x_1)$

From earlier results:
- $Var(X_2|x_1) = E(X_2^2|x_1) - (E(X_2|x_1))^2$
- $E(u(X_1)|x_2) = \int_{-\infty}^{\infty} u(x_1) f_{1|2}(x_1|x_2)\,dx_1$

Finally: Let (X_1, X_2) be a random vector such that $Var(X_2)$ is finite.
↳ $E(E(X_2|X_1)) = E(X_2)$ * useful
↳ $Var(E(X_2|X_1)) \le Var(X_2)$

General bivariate transformation: We start w/ (X_1, X_2) and want (Y_1, Y_2), where both are functions of $(X_1$ and $X_2)$. We express X_1 and X_2 in terms of Y_1 and Y_2 (get inverses). Now, $P((Y_1,Y_2) \in A) = \iint f_{X_1X_2}(x_1,x_2)\,dx_1 dx_2 = \iint_B f_{X_1X_2}(w_1(y_1,y_2), w_2(y_1,y_2))|J|\,dy_1 dy_2$ ↳ plug in original bounds for new bounds ↳ pdf is integrand. Perform integrals.

$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$

Series calculations: n terms in series (a_1 to a_n); S_n = first n terms summed; d = diff between succ. terms
Arithmetic: $a_n = a_1 + (n-1)d$; $a_i = \dfrac{a_{i-1} + a_{i+1}}{2}$
$S_n = \dfrac{a_1 + a_n}{2} \cdot n$; $S_n = \dfrac{2a_1 + (n-1)d}{2} \cdot n$
Geometric: $a_n = a_1 \cdot q^{n-1}$; $a_i = \sqrt{a_{i-1} \cdot a_{i+1}}$
$S_n = \dfrac{a_n q - a_1}{q - 1}$; $S_n = \dfrac{a_1(q^n - 1)}{q - 1}$

Powers of natural #s:
$\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$
$\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$
$\sum_{k=1}^{n} k^3 = \frac{1}{4}n^2(n+1)^2$

Special power series: $\frac{1}{1-x} = 1 + x + x^2 + \dots$
$\frac{1}{x+1} = 1 - x + x^2 - x^3 + \dots$; $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$
$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$; $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$
$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$; $\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \dots$

Taylor and Maclaurin:
$f(x) = f(a) + f'(a)(x-a) + \dfrac{f''(a)(x-a)^2}{2!} + \dots + \dfrac{f^{(n-1)}(a)(x-a)^{n-1}}{(n-1)!} + R_n$; $R_n = \dfrac{f^{(n)}(\xi)(x-a)^n}{n!}$, $a \le \xi \le x$

Binomial series:
$(a+x)^n = a^n + \binom{n}{1}a^{n-1}x + \binom{n}{2}a^{n-2}x^2 + \binom{n}{3}a^{n-3}x^3 + \dots$

Let X have pmf $p_X(x) = \begin{cases} \frac{3!}{x!(3-x)!}(2/3)^x(1/3)^{3-x}, & x = 0,1,2,3 \\ 0, & \text{elsewhere} \end{cases}$
We seek pmf $p_Y(y)$ of r.v. $Y = X^2$. Transformation $y = g(x) = x^2$ maps onto $y = 0, 1, 4, 9$. In general, not a one-to-one, but here yes. $x = \sqrt{y}$. So $p_Y(y) = p_X(\sqrt{y}) = \dfrac{3!}{\sqrt{y}!(3-\sqrt{y})!}(2/3)^{\sqrt{y}}(1/3)^{3-\sqrt{y}}$ for $y = 0,1,4,9$; 0 elsewhere.

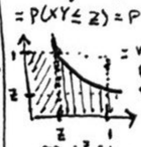
1st head appears odd, pay \$1, if even, win \$1. Space Y is {1, −1}. Dist. given by $p_Y(-1) = 2/3$, $p_Y(1) = 1/3$. X = # flips until heads appears. odd = 2/3.
Now if Z is $(X-2)^2$ from game above, not 1-to-1, so we split up and treat each unique case differently.
$p_Z(z) = \begin{cases} p_X(2) = 1/4 & \text{for } z = 0 \\ p_X(1) + p_X(3) = 5/8 & \text{for } z = 1 \\ p_X(\sqrt{z}+2) = \frac{1}{4}(\frac{1}{2})^{\sqrt{z}} & \text{for } z = 4, 9, 16\dots \end{cases}$

A lot has 100 fuses, five to be tested. If all 5 are good, lot accepted. pmf for X = # defective inspected. Assume 20 actually defective.
$p_X(x) = \begin{cases} \dfrac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}} & \text{for } x = 1,2,3,4,5 \\ 0 & \text{elsewhere} \end{cases}$

Derivatives: $\frac{d}{dx}(cf(x)) = cf'(x)$
$(f(x) \pm g(x))' = f'(x) \pm g'(x)$
product $(fg)' = f'g + fg'$; quotient $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$
$\frac{d}{dx}(x^n) = nx^{n-1}$; $\frac{d}{dx}(c) = 0$
$\frac{d}{dx}(f(g(x))) = f'(g(x))g'(x)$
$\frac{d}{dx}\sin x = \cos x$; $\frac{d}{dx}\cos x = -\sin x$; $\frac{d}{dx}\tan x = \sec^2 x$
$\frac{d}{dx}\sec x = \sec x \tan x$; $\frac{d}{dx}\csc x = -\csc x \cot x$; $\frac{d}{dx}\cot x = -\csc^2 x$
$\frac{d}{dx}\sin^{-1}x = \frac{1}{\sqrt{1-x^2}}$; $\frac{d}{dx}\cos^{-1}x = \frac{-1}{\sqrt{1-x^2}}$; $\frac{d}{dx}\tan^{-1}x = \frac{1}{1+x^2}$
$\frac{d}{dx}\sec^{-1}x = \frac{1}{|x|\sqrt{x^2-1}}$; $\frac{d}{dx}\csc^{-1}x = \frac{-1}{|x|\sqrt{x^2-1}}$; $\cot^{-1}x = \frac{-1}{1+x^2}$
$\frac{d}{dx}(a^x) = a^x \ln(a)$; $\frac{d}{dx}e^x = e^x$
$\frac{d}{dx}(\ln(x)) = \frac{1}{x}, x > 0$; $\frac{d}{dx}(\ln|x|) = \frac{1}{x}, x \ne 0$
$\frac{d}{dx}(\log_a(x)) = \frac{1}{x \ln a}, x > 0$

Random point in unit circle:
X = distance to origin
Probability of point lying in a smaller circle C is $P(C) = \frac{\text{area } C}{\pi}$.
For $0 < x < 1$, event $\{X \le x\}$ is point lying in circle of radius x.
So $P(X \le x) = \frac{\pi x^2}{\pi} = x^2$, so cdf is $F_X(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \le x < 1 \\ 1, & 1 \le x \end{cases}$
pdf is $f_X(x) = \begin{cases} 2x, & 0 \le x < 1 \\ 0, & \text{elsewhere} \end{cases}$
$P(\frac{1}{4} < X \le \frac{1}{2}) = \int_{1/4}^{1/2} 2w\,dw = \frac{3}{16}$

$f(x,y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$
$Z = XY$; $F(z) = P(Z \le z)$
$= P(XY \le z) = P(Y \le \frac{z}{x})$

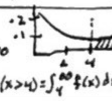


= volume under $f(x,y)$ over shaded area
$= \int_0^z \int_0^1 1\,dy\,dx + \int_z^1 \int_0^{z/x} 1\,dy\,dx$
$f(z) = \frac{d}{dz}F(z)$

$f(x) = \begin{cases} 1/3, & -1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$
$Y = X^2$; $F(Y) = P(Y \le y) = P(X^2 \le y)$
Two cases: ① $0 \le y < 1$; ② $1 \le y < 4$
① $P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y})$
$= \int_{-1}^{\sqrt{y}} 1/3\,dx - \int_{-1}^{-\sqrt{y}} 1/3\,dx = \frac{2\sqrt{y}}{3}$
② $P(X^2 \le y) = P(X \le \sqrt{y}) = \int_{-1}^{\sqrt{y}} 1/3\,dx = \frac{1+\sqrt{y}}{3}$
$F(y) = \begin{cases} 0, & -\infty < y < 0 \\ 2\sqrt{y}/3, & 0 < y < 1 \\ (1+\sqrt{y})/3, & 1 < y < 4 \\ 1, & y \ge 4 \end{cases}$
$f(y) = \frac{d}{dy}F(y)$
$F(Y) = \begin{cases} \frac{1}{3\sqrt{y}}, & 0 < y < 1 \\ \frac{1}{6\sqrt{y}}, & 1 < y < 4 \\ 0, & \text{elsewhere} \end{cases}$

$f(x) = \begin{cases} x^2/9, & 0 \le x < 3 \\ 0, & \text{elsewhere} \end{cases}$ $Y = X^3$
$F(Y) = P(Y \le y) = P(X^3 \le y)$
$= P(X \le \sqrt[3]{y}) = \int_0^{\sqrt[3]{y}} x^2/9\,dx$
$f(y) = \frac{d}{dy}F(Y)$ Y support is now $0 < y < 27$

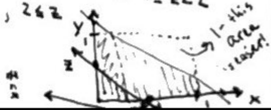
Calls to switchboard
$f(x) = \begin{cases} \frac{e^{-x/4}}{4}, & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$



$P(X > 4) = \int_4^{\infty} f(x)\,dx$

$p_{X_1X_2}(x_1,x_2) = \dfrac{\mu_1^{x_1}\mu_2^{x_2}e^{-\mu_1}e^{-\mu_2}}{x_1!\,x_2!}$ pmf, x_1 and x_2 nonneg. integers
We want pmf of $Y = X_1 + X_2$. We need to define a Y_2, let's choose $Y_2 = X_2$.
Inverse functions are $x_1 = y_1 - y_2$ and $x_2 = y_2$. Thus, $p_{Y_1Y_2}(y_1,y_2) = \dfrac{\mu_1^{y_1-y_2}\mu_2^{y_2}e^{-\mu_1-\mu_2}}{(y_1-y_2)!\,y_2!}$
*Note: $y_2 = 0, 1, \dots, y_1$; $y_1 = 0, 1, 2, \dots$

Random point on unit square. We want not X or Y, but $Z = X+Y$. $f_{XY}(x,y) = \begin{cases} 1, & \text{for...} \\ 0, & \text{elsewhere} \end{cases}$
$F_Z(z) = P(X + Y \le z) = \begin{cases} 0, & z < 0 \\ \int_0^z \int_0^{z-x} dy\,dx, & 0 < z < 1 \\ 1 - \int_{z-1}^1 \int_{z-x}^1 dy\,dx & \text{for } 1 \le z < 2 \\ 1, & 2 \le z \end{cases}$



Integrals:
$\int x^{-n}\,dx = \left(\frac{1}{-n+1}\right)x^{-n+1}, n \ne 1$
$\int \frac{1}{ax+b}\,dx = \frac{1}{a}\ln|ax+b|$; $\int x^{p/q}\,dx = \frac{q}{p+q}x^{\frac{p+q}{q}}$
$\int e^u\,du = e^u$; $\int a^u\,du = \frac{a^u}{\ln a}$; $\int \ln u\,du = u\ln u - u$
$\int ue^u\,du = (u-1)e^u$; $\int \frac{1}{u \ln u}\,du = \ln|\ln u|$

Whenever you divide or multiply by a negative number, you must flip the inequality sign!

$A \cup B = B \cup A$

$A \cap B = B \cap A$

Associativity:

$A \cup (B \cup C) = (A \cup B) \cup C$

$A \cap (B \cap C) = (A \cap B) \cap C$

Distributive laws:

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

DeMorgan's laws:

$(A \cup B)^c = A^c \cap B^c$

$(A \cap B)^c = A^c \cup B^c$

Basic properties of $P(\cdot)$:

① $P(\emptyset) = 0$

② $0 \le P(A)$

③ $P(A^c) = 1 - P(A)$

④ If $A_1 \subset A_2$, then $P(A_1) \le P(A_2)$

⑤ $P(B \cap A^c) = P(B) - P(A \cap B)$

⑥ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

↳ $P(B \cup A) \le P(A) + P(B)$ "Boole's"

↳ $P(A^c \cap B^c) \ge 1 - (P(A) + P(B))$

Conditional probability and independence:

Def'n: $P(A \cap B) = P(A)P(B)$ means statistical independence

↳ from Bayes' w/ assumption that given has no effect

↳ If A and B are independent, so are

- A and B^c
- A^c and B
- A^c and B^c

Def'n conditional prob: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

↳ $P(A \cap B) = P(A|B)P(B)$

↳ $P(A \cap B) = P(B|A)P(A)$

↳ $P(A|B) = P(B|A)\left(\frac{P(A)}{P(B)}\right)$ "Bayes"

$\binom{5}{0}\binom{2}{2}$ ... $\binom{52}{5}$

	with replacement	without replacement
ordered	n^r	$\frac{n!}{(n-r)!}$
unordered	$\binom{n+r-1}{r}$	$\binom{n}{r} = \frac{n!}{r!(n-r)!}$

cdf of X defined by $F_X(x) = P(X \le x)$ for all x

$F_X(x)$ is a cdf iff

a) $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$

b) $F_X(x)$ is a nondecreasing function of x

c) $F_X(x)$ is right-continuous: for every x_0, $\lim_{x \downarrow x_0} F_X(x) = F(x_0)$

The following two statements are equivalent:

1) r.v.s X and Y are identically distributed

2) $F_X(x) = F_Y(x)$ for every x

pmf of a discrete r.v. X given by $f_X(x) = P(X = x)$ for all x

pdf of a continuous r.v. X is the function that satisfies $F_X(x) = \int_{-\infty}^x f_X(t)\,dt$ for all x

$P(X \le x) = F_X(x) = \int_{-\infty}^x f_X(t)\,dt$ for all x

Thus, $\frac{d}{dx}F_X(x) = f_X(x)$

$f_X(x)$ is a pdf or pmf iff:

a) $f_X(x) \ge 0$ for all x

b) $\sum_x f_X(x) = 1$ (pmf) OR $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ (pdf)

Expectation:

$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$ OR $E(X) = \sum_x x f(x)$

$E(k_1 g_1(x) + k_2 g_2(x) + c) = k_1 E(g_1(x)) + k_2 E(g_2(x)) + c$

LOTUS: Suppose X is a continuous r.v. defined by pdf f(x). Let Y be a r.v. $Y = g(X)$ for some functiong.

$E(Y) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

Suppose (X_1, X_2) is a continuous random vector. Let Y be a r.v. defined by $Y = g(X_1, X_2)$ for some real-valued function g (i.e. $\mathbb{R}^2 \to \mathbb{R}$). Then $E(Y) = \iint_{-\infty}^{\infty} g(x_1, x_2) f(x_1, x_2)\,dx_1 dx_2$ (↳ joint pdf)

$E(k_1 Y_1 + k_2 Y_2) = k_1 E(Y_1) + k_2 E(Y_2)$

Variance: Let X be a r.v. with finite mean $\mu = E(X)$ and such that $E((X-\mu)^2)$ is finite. Then the variance of X is: $\sigma^2 = E((X-\mu)^2)$

$= E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2$

$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$

Multivariate:

Joint cdf: $P(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2) = F_{X_1X_2}(b_1, b_2) - F_{X_1X_2}(a_1, b_2) - F_{X_1X_2}(b_1, a_2) + F_{X_1X_2}(a_1, a_2)$

Joint pmf: $p_{X_1X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$

Joint pdf: $F_{X_1X_2}(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{X_1X_2}(w_1, w_2)\,dw_1 dw_2$

$\frac{\partial^2 F_{X_1X_2}(x_1, x_2)}{\partial x_1 \partial x_2} = f_{X_1X_2}(x_1, x_2)$

Note that $P((X_1, X_2) \in A) = \iint_A f_{X_1X_2}(x_1, x_2)\,dx_1 dx_2$, which is just the volume under the surface $z = f_{X_1X_2}(x_1, x_2)$ over the set A.

Marginals: To find marginal pmfs, hold marginal variable and sum over the other.

To find marginal pdfs, $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\,dx_2$ ↳ joint w/ respect to other r.v.

To find marginal cdfs, $F_{X_1}(x_1) = P(X_1 \le x_1, -\infty < X_2 < \infty) = \lim_{x_2 \to \infty} F(x_1, x_2)$

Conditional distributions and expect.:

Discrete: $P(X_2 = x_2 | X_1 = x_1) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)}$

$= \frac{p_{X_1X_2}(x_1, x_2)}{p_{X_1}(x_1)} = p_{X_2|X_1}(x_2|x_1)$ → conditional pmf

Continuous: $f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_1}(x_1)}$

Note: $\int_{-\infty}^{\infty} f_{X_2|X_1}(x_2|x_1)\,dx_2 = \int_{-\infty}^{\infty} \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_1}(x_1)}\,dx_2$

$= \frac{1}{f_{X_1}(x_1)}\int_{-\infty}^{\infty} f_{X_1X_2}(x_1, x_2)\,dx_2 = \frac{f_{X_1}(x_1)}{f_{X_1}(x_1)} = 1$

$P(a < X_2 < b | X_1 = x_1) = \int_a^b f_{2|1}(x_2|x_1)\,dx_2$

$P(c < X_1 < d | X_2 = x_2) = \int_c^d f_{1|2}(x_1|x_2)\,dx_1$

If $u(X_2)$ is a function of X_2, the conditional expectation of $u(X_2)$, given that $X_1 = x_1$, is

$E(u(X_2)|x_1) = \int_{-\infty}^{\infty} u(x_2) f_{2|1}(x_2|x_1)\,dx_2$

Conditional variance: $E((X_2 - E(X_2|x_1))^2 | x_1)$

Conditional mean: $E(X_2|x_1)$

From earlier results:

- $Var(X_2|x_1) = E(X_2^2|x_1) - (E(X_2|x_1))^2$
- $E(u(X_1)|x_2) = \int_{-\infty}^{\infty} u(x_1) f_{1|2}(x_1|x_2)\,dx_1$

Finally: Let (X_1, X_2) be a random vector such that $Var(X_2)$ is finite.

↳ $E(E(X_2|X_1)) = E(X_2)$ * useful

↳ $Var(E(X_2|X_1)) \le Var(X_2)$

General bivariate transformation: We start w/ (X_1, X_2) and want (Y_1, Y_2), where both are functions of X_1 and X_2. We express X_1 and X_2 in terms of Y_1 and Y_2 (find inverses), then $P((Y_1, Y_2) \in B) = \iint_B f_{X_1X_2}(w_1(y_1,y_2), w_2(y_1,y_2))|J|\,dy_1 dy_2$ ... $P((X_1,X_2)\in A) = \iint_A f_{X_1X_2}(x_1,x_2)\,dx_1dx_2$ ... with new bounds ... pdf is integrand

$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$

Series calculations: n terms in series (a_1 to a_n); S_n = first n terms summed; d = diff between succ. terms

Arithmetic: $a_n = a_1 + (n-1)d$; $a_i = \frac{a_{i-1} + a_{i+1}}{2}$

$S_n = \frac{a_1 + a_n}{2} \cdot n$; $S_n = \frac{2a_1 + (n-1)d}{2} \cdot n$

Geometric: $a_n = a_1 \cdot q^{n-1}$; $a_i = \sqrt{a_{i-1} \cdot a_{i+1}}$

$S_n = \frac{a_n q - a_1}{q - 1}$; $S_n = \frac{a_1(q^n - 1)}{q - 1}$

Powers of natural #s:

$\sum_{k=1}^n k = \frac{1}{2}n(n+1)$

$\sum_{k=1}^n k^2 = \frac{1}{6}n(n+1)(2n+1)$

$\sum_{k=1}^n k^3 = \frac{1}{4}n^2(n+1)^2$

Special power series: $\frac{1}{1-x} = 1 + x + x^2 + \ldots$

$\frac{1}{x+1} = 1 - x + x^2 - x^3 + \ldots$; $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$

$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$; $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots$

$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots$; $\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \ldots$

Taylor and Maclaurin:

$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \ldots + \frac{f^{(n-1)}(a)(x-a)^{n-1}}{(n-1)!} + R_n$

$R_n = \frac{f^{(n)}(\xi)(x-a)^n}{n!}$, $a \le \xi \le x$

Binomial series:

$(a+x)^n = a^n + \binom{n}{1}a^{n-1}x + \binom{n}{2}a^{n-2}x^2 + \binom{n}{3}a^{n-3}x^3 + \ldots$

Let X have pmf $p_X(x) = \begin{cases} \frac{3!}{x!(3-x)!}(2/3)^x(1/3)^{3-x}, & x = 0, 1, 2, 3 \\ 0, & \text{elsewhere} \end{cases}$

We seek pmf $p_Y(y)$ of r.v. $Y = X^2$. Transformation $y = g(x) = x^2$ maps onto $y = 0, 1, 4, 9$. In general, not a one-to-one, but here yes. $x = \sqrt{y}$. So $p_Y(y) = p_X(\sqrt{y}) = \frac{3!}{\sqrt{y}!(3-\sqrt{y})!}(2/3)^{\sqrt{y}}(1/3)^{3-\sqrt{y}}$ for $y = 0, 1, 4, 9$; 0 elsewhere.

1st head appears odd, pay #1, if even, win #1. Space Y is {-1, 1}. Dist. given by $p_Y(-1) = 1/3$, $p_Y(1) = 2/3$.

Now if Z is $(X-2)^2$ from game above, X = # turns tossed, not 1-to-1, so we split up and treat each unique case differently.

X = # flips until heads appears, odd = 2/3

$p_Z(z) = \begin{cases} p_X(2) = 1/4 & \text{for } z = 0 \\ p_X(1) + p_X(3) = 5/8 & \text{for } z = 1 \\ p_X(\sqrt{z} + 2) = \frac{1}{2}(\frac{1}{2})^{\sqrt{z}} & \text{for } z = 4, 9, 16\ldots \end{cases}$

A lot has 100 fuses, five to be tested. If all 5 are good, lot accepted. pmf for X = # defects inspected. Assume 20 are actually defective.

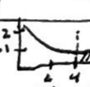
$p_X(x) = \begin{cases} \frac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}} \\ \text{for } x = 1, 2, 3, 4, 5 \\ 0 \text{ elsewhere} \end{cases}$

Random point in unit circle: X = distance to origin. Probability of point lying in a smaller circle C is $P(C) = \frac{area\ C}{\pi}$. For $0 < x < 1$, event $\{X \le x\}$ is point lying in circle of radius x. So $P(X \le x) = \frac{\pi x^2}{\pi} = x^2$, so cdf is $F_X(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \le x < 1 \\ 1, & 1 \le x \end{cases}$

pdf is $f_X(x) = \begin{cases} 2x, & 0 \le x < 1 \\ 0, & \text{elsewhere} \end{cases}$

$P(\frac{1}{4} < X \le \frac{1}{2}) = \int_{1/4}^{1/2} 2w\,dw = \frac{3}{16}$

Calls to switchboard

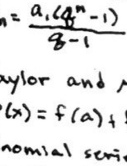


$f(x) = \begin{cases} \frac{e^{-x/4}}{4}, & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$ $P(X > 4) = \int_4^\infty f(x)\,dx$

$f(x, y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$

$Z = XY$; $F(z) = P(Z \le z)$

$= P(XY \le z) = P(Y \le \frac{z}{x})$



= volume under $f(x,y)$ over shaded area

$= \int_0^z \int_0^1 1\,dy\,dx + \int_z^1 \int_0^{z/x} 1\,dy\,dx$

$f(z) = \frac{d}{dz}F(z)$

$f(x) = \begin{cases} x^2/9, & 0 \le x < 3 \\ 0, & \text{elsewhere} \end{cases}$ $Y = X^3$

$F(Y) = P(Y \le y) = P(X^3 \le y)$

$= P(X \le \sqrt[3]{y}) = \int_0^{\sqrt[3]{y}} x^2/9\,dx$

$f(y) = \frac{d}{dy}F(Y)$ Y support is now $0 \le y < 27$

$f(x) = \begin{cases} 1/3, & -1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$

$Y = X^2$; $F(Y) = P(Y \le y) = P(X^2 \le y)$

Two cases: ① $0 \le y < 1$; ② $1 \le y < 4$

① $P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y})$

$= \int_{-1}^{\sqrt{y}} 1/3\,dx - \int_{-1}^{-\sqrt{y}} 1/3\,dx = \frac{2\sqrt{y}}{3}$

② $P(X^2 \le y) = P(X \le \sqrt{y}) = \int_{-1}^{\sqrt{y}} 1/3\,dx$

$= \frac{1 + \sqrt{y}}{3}$

$F(y) = \begin{cases} 0, & -\infty < y < 0 \\ 2\sqrt{y}/3, & 0 \le y < 1 \\ (1+\sqrt{y})/3, & 1 \le y < 4 \\ 1, & y \ge 4 \end{cases}$

$f(y) = \frac{d}{dy}F(y)$

$f(Y) = \begin{cases} \frac{1}{3\sqrt{y}}, & 0 < y < 1 \\ \frac{1}{6\sqrt{y}}, & 1 \le y < 4 \\ 0, & \text{elsewhere} \end{cases}$

$p_{X_1X_2}(x_1, x_2) = \frac{\mu_1^{x_1}\mu_2^{x_2}e^{-\mu_1}e^{-\mu_2}}{x_1!\,x_2!}$ pmf, x_1 and x_2 nonneg. integers

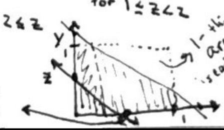
We want pmf of $Y_1 = X_1 + X_2$. We need to define a Y_2, let's choose $Y_2 = X_2$.

Inverse functions are $x_1 = y_1 - y_2$ and $x_2 = y_2$. Thus, $p_{Y_1Y_2}(y_1, y_2) = \frac{\mu_1^{y_1-y_2}\mu_2^{y_2}e^{-\mu_1-\mu_2}}{(y_1-y_2)!\,y_2!}$

*Note: $y_2 = 0, 1, \ldots, y_1$; $y_1 = 0, 1, 2, \ldots$

Random point on unit square. We want not X or Y, but $Z = X + Y$. $f_{XY}(x,y) = \begin{cases} 1, & \text{on square} \\ 0, & \text{elsewhere} \end{cases}$

$F_Z(z) = P(X + Y \le z) = \begin{cases} 0, & z < 0 \\ \int_0^z\int_0^{z-x} dy\,dx, & 0 \le z < 1 \text{ (make line)} \\ 1 - \int_{z-1}^1\int_{z-x}^1 dy\,dx & \text{for } 1 \le z < 2 \\ 1, & 2 \le z \end{cases}$



Derivatives: $\frac{d}{dx}(cf(x)) = cf'(x)$

$(f(x) \pm g(x))' = f'(x) \pm g'(x)$

product: $(f \cdot g)' = f'g + fg'$; quotient: $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$

$\frac{d}{dx}(x^n) = nx^{n-1}$; $\frac{d}{dx}(c) = 0$

$\frac{d}{dx}(f(g(x))) = f'(g(x))g'(x)$

$\frac{d}{dx}\sin x = \cos x$; $\frac{d}{dx}\cos x = -\sin x$; $\frac{d}{dx}\tan x = \sec^2 x$

$\frac{d}{dx}\sec x = \sec x \tan x$; $\frac{d}{dx}\csc x = -\csc x \cot x$; $\frac{d}{dx}\cot x = -\csc^2 x$

$\frac{d}{dx}\sin^{-1}x = \frac{1}{\sqrt{1-x^2}}$; $\frac{d}{dx}\cos^{-1}x = \frac{-1}{\sqrt{1-x^2}}$; $\frac{d}{dx}\tan^{-1}x = \frac{1}{1+x^2}$

$\frac{d}{dx}\sec^{-1}x = \frac{1}{|x|\sqrt{x^2-1}}$; $\frac{d}{dx}\csc^{-1}x = \frac{-1}{|x|\sqrt{x^2-1}}$; $\cot^{-1}x = \frac{-1}{1+x^2}$

$\frac{d}{dx}(a^x) = a^x \ln(a)$; $\frac{d}{dx}e^x = e^x$

$\frac{d}{dx}(\ln(x)) = \frac{1}{x}$, $x > 0$; $\frac{d}{dx}(\ln|x|) = \frac{1}{x}$, $x \ne 0$

$\frac{d}{dx}(\log_a(x)) = \frac{1}{x \ln a}$, $x > 0$

Integrals:

$\int x^{-n}\,dx = \left(\frac{1}{-n+1}\right)x^{-n+1}$, $n \ne 1$

$\int \frac{1}{ax+b}\,dx = \frac{1}{a}\ln|ax+b|$; $\int x^{p/q}\,dx = \frac{q}{p+q}x^{\frac{p+q}{q}}$

$\int e^u\,du = e^u$; $\int a^u\,du = \frac{a^u}{\ln a}$; $\int \ln u\,du = u\ln u - u$

$\int ue^u\,du = (u-1)e^u$; $\int \frac{1}{u\ln u}\,du = \ln|\ln u|$

Whenever you divide or multiply by a negative number, you must flip the inequality sign!

Commutativity:
$A \cup B = B \cup A$
$A \cap B = B \cap A$

Associativity:
$A \cup (B \cup C) = (A \cup B) \cup C$
$A \cap (B \cap C) = (A \cap B) \cap C$

Distributive laws:
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

DeMorgan's laws:
$(A \cup B)^c = A^c \cap B^c$
$(A \cap B)^c = A^c \cup B^c$

$P(\text{full house}) = \dfrac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}$

	ordered	unordered
with replacement	n^r	$\binom{n+r-1}{r}$
without replacement	$\frac{n!}{(n-r)!}$	$\binom{n}{r}$

Basic properties of $P(\cdot)$:
① $P(\emptyset) = 0$
② $0 \le P(A)$
③ $P(A^c) = 1 - P(A)$
④ If $A_1 \subset A_2$, then $P(A_1) \le P(A_2)$
⑤ $P(B \cap A^c) = P(B) - P(A \cap B)$
⑥ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
↳ $P(B \cup A) \le P(A) + P(B)$ "Boole's"
↳ $P(A^c \cap B^c) \ge 1 - (P(A) + P(B))$

Conditional probability and independence:
Def'n: $P(A \cap B) = P(A)P(B)$ means statistical independence
↳ from Bayes' w/ assumption that given has no effect
↳ If A and B are independent, so are
- A and B^c
- A^c and B
- A^c and B^c

Def'n conditional prob: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$
↳ $P(A \cap B) = P(A|B)P(B)$
↳ $P(A \cap B) = P(B|A)P(A)$
↳ $P(A|B) = P(B|A)\left(\dfrac{P(A)}{P(B)}\right)$ "Bayes"

cdf of X defined by $F_X(x) = P(X \le x)$ for all x
$F_X(x)$ is a cdf iff
a) $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$
b) $F_X(x)$ is a nondecreasing function of x
c) $F_X(x)$ is right-continuous: for every x_0, $\lim_{x\downarrow x_0} F_X(x) = F(x_0)$

The following two statements are equivalent:
1) r.v.s X and Y are identically distributed
2) $F_X(x) = F_Y(x)$ for every x

pmf of a discrete r.v. X given by $f_X(x) = P(X=x)$ for all x

pdf of a continuous r.v. X is the function that satisfies $F_X(x) = \int_{-\infty}^x f_X(t)\,dt$ for all x
$P(X \le x) = F_X(x) = \int_{-\infty}^x f_X(t)\,dt$ for all x
Thus, $\frac{d}{dx}F_X(x) = f_X(x)$

$f_X(x)$ is a pdf or pmf iff:
a) $f_X(x) \ge 0$ for all x
b) $\sum_x f_X(x) = 1$ (pmf) OR $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ (pdf)

Multivariate:
Joint cdf: $P(a_1 < X_1 \le b_1,\ a_2 < X_2 \le b_2) = F_{X_1X_2}(b_1,b_2) - F_{X_1X_2}(b_1,a_2) - F_{X_1X_2}(a_1,b_2) + F_{X_1X_2}(a_1,a_2)$

Joint pmf: $p_{X_1,X_2}(x_1,x_2) = P(X_1 = x_1, X_2 = x_2)$

Joint pdf: $F_{X_1X_2}(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{X_1X_2}(w_1,w_2)\,dw_1\,dw_2$

$\dfrac{\partial^2 F_{X_1X_2}(x_1,x_2)}{\partial x_1 \partial x_2} = f_{X_1X_2}(x_1,x_2)$

Note that $P((X_1,X_2) \in A) = \iint_A f_{X_1X_2}(x_1,x_2)\,dx_1dx_2$, which is just the volume under the surface $z = f_{X_1X_2}(x_1,x_2)$ over the set A.

Marginals: To find marginal pmfs, hold marginal variable and sum over the other.
To find marginal pdfs, $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1,x_2)\,dx_2$ ↗ violate w/ respect to other r.v.
To find marginal cdfs, $F_{X_1}(x_1) = P(X_1 \le x_1, -\infty < X_2 < \infty) = \lim_{x_2\to\infty} F(x_1,x_2)$

Conditional distributions and expect.:
Discrete: $P(X_2 = x_2 | X_1 = x_1) = \dfrac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)} = \dfrac{p_{X_1X_2}(x_1,x_2)}{p_{X_1}(x_1)} = p_{X_2|X_1}(x_2|x_1)$ ← conditional pmf

Continuous: $f_{X_2|X_1}(x_2|x_1) = \dfrac{f_{X_1X_2}(x_1,x_2)}{f_{X_1}(x_1)}$

Note: $\int_{-\infty}^{\infty} f_{X_2|X_1}(x_2|x_1)\,dx_2 = \int_{-\infty}^{\infty} \dfrac{f_{X_1X_2}(x_1,x_2)}{f_{X_1}(x_1)}\,dx_2 = \dfrac{1}{f_{X_1}(x_1)}\int_{-\infty}^{\infty} f_{X_1X_2}(x_1,x_2)\,dx_2 = \dfrac{f_{X_1}(x_1)}{f_{X_1}(x_1)} = 1$

$P(a < X_2 < b | X_1 = x_1) = \int_a^b f_{2|1}(x_2|x_1)\,dx_2$
$P(c < X_1 < d | X_2 = x_2) = \int_c^d f_{1|2}(x_1|x_2)\,dx_1$

If $u(X_2)$ is a function of X_2, the conditional expectation of $u(X_2)$, given that $X_1 = x_1$, is
$E(u(X_2)|x_1) = \int_{-\infty}^{\infty} u(x_2) f_{2|1}(x_2|x_1)\,dx_2$
Conditional variance: $E((X_2 - E(X_2|x_1))^2|x_1)$
Conditional mean: $E(X_2|x_1)$

From earlier results:
- $Var(X_2|x_1) = E(X_2^2|x_1) - (E(X_2|x_1))^2$
- $E(u(X_1)|x_2) = \int_{-\infty}^{\infty} u(x_1) f_{1|2}(x_1|x_2)\,dx_1$

Finally: Let (X_1,X_2) be a random vector such that $Var(X_2)$ is finite.
↳ $E(E(X_2|X_1)) = E(X_2)$ * useful
↳ $Var(E(X_2|X_1)) \le Var(X_2)$

General bivariate transformation: We start w/ (X_1,X_2) and want (Y_1,Y_2), where both are functions of $(X_1$ and $X_2)$. We express X_1 and X_2 in terms of Y_1 and Y_2 (inverse functions). $P((Y_1,Y_2)\in B) = \iint_B f_{X_1X_2}(w_1(y_1,y_2), w_2(y_1,y_2))\,|J|\,dy_1dy_2$ [illegible]

Expectation:
$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$ OR $E(X) = \sum_x x f(x)$
$E(k_1 g_1(X) + k_2 g_2(X) + c) = k_1 E(g_1(X)) + k_2 E(g_2(X)) + c$

LOTUS: Suppose X is a continuous r.v. defined by pdf f(x). Let Y be a r.v. $Y = g(X)$ for some function g.
$E(Y) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

Suppose (X_1,X_2) is a continuous random vector. Let Y be a r.v. defined by $Y = g(X_1,X_2)$ for some real-valued function g (i.e. $\mathbb{R}^2 \to \mathbb{R}$). Then $E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1,x_2) f(x_1,x_2)\,dx_1dx_2$ ← joint pdf

$E(k_1Y_1 + k_2Y_2) = k_1E(Y_1) + k_2E(Y_2)$

Variance: Let X be a r.v. with finite mean $\mu = E(X)$ and such that $E((X-\mu)^2)$ is finite. Then the variance of X is: $\sigma^2 = E((X-\mu)^2)$
$= E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2$
$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$

Series calculations: n terms in series (a_1 to a_n); S_n = first n terms summed; d = diff between succ. terms

Arithmetic: $a_n = a_1 + (n-1)d$; $a_i = \dfrac{a_{i-1} + a_{i+1}}{2}$
$S_n = \dfrac{a_1 + a_n}{2}\cdot n$; $S_n = \dfrac{2a_1 + (n-1)d}{2}\cdot n$

Geometric: $a_n = a_1 \cdot q^{n-1}$; $a_i = \sqrt{a_{i-1}\cdot a_{i+1}}$
$S_n = \dfrac{a_n q - a_1}{q-1}$; $S_n = \dfrac{a_1(q^n - 1)}{q-1}$

Powers of natural #s:
$\sum_{k=1}^n k = \frac{1}{2}n(n+1)$
$\sum_{k=1}^n k^2 = \frac{1}{6}n(n+1)(2n+1)$
$\sum_{k=1}^n k^3 = \frac{1}{4}n^2(n+1)^2$

Special power series: $\frac{1}{1-x} = 1 + x + x^2 + \dots$
$\frac{1}{x+1} = 1 - x + x^2 - x^3 + \dots$; $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$
$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$; $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$
$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$; $\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \dots$

Taylor and Maclaurin:
$f(x) = f(a) + f'(a)(x-a) + \dfrac{f''(a)(x-a)^2}{2!} + \dots + \dfrac{f^{(n-1)}(a)(x-a)^{n-1}}{(n-1)!} + R_n$; $R_n = \dfrac{f^{(n)}(\xi)(x-a)^n}{n!}$, $a \le \xi \le x$

Binomial series:
$(a+x)^n = a^n + \binom{n}{1}a^{n-1}x + \binom{n}{2}a^{n-2}x^2 + \binom{n}{3}a^{n-3}x^3 + \dots$

Let X have pmf $p_X(x) = \begin{cases} \frac{3!}{x!(3-x)!}(2/3)^x(1/3)^{3-x}, & x = 0,1,2,3 \\ 0, & \text{elsewhere} \end{cases}$
We seek pmf $p_Y(y)$ of r.v. $Y = X^2$. Transformation $y = g(x) = x^2$ maps onto $y = 0, 1, 4, 9$. In general, not a one-to-one, but here yes. $x = \sqrt{y}$.
So $p_Y(y) = p_X(\sqrt{y}) = \dfrac{3!}{\sqrt{y}!(3-\sqrt{y})!}(2/3)^{\sqrt{y}}(1/3)^{3-\sqrt{y}}$ for $y = 0,1,4,9$; 0 elsewhere.

A lot has 100 fuses, five to be tested. If all 5 are good, lot accepted. pmf for X = # defective inspected. Assume 20 are already defective.
$p_X(x) = \dfrac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}}$ for $x = 1,2,3,4,5$; 0 elsewhere.

1st head appears odd, pay #1, if even, win #1. Space Y is {1, -1}.
Now if Z is $(X-2)^2$ from same above, not 1-to-1, so we split up and treat each unique case differently. X = # flips until heads appears. odd = 2/3. Dist. given by $p_Y(-1) = 1/3$, $p_Y(1) = 2/3$
$P_Z(z) = \begin{cases} p_X(2) = 1/4 & \text{for } z = 0 \\ p_X(1) + p_X(3) = 5/8 & \text{for } z = 1 \\ p_X(\sqrt{z}+2) = \frac{1}{2}(\frac{1}{2})^{\sqrt{z}} & \text{for } z = 4, 9, 16\dots \end{cases}$

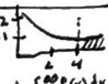
Derivatives: $\frac{d}{dx}(cf(x)) = cf'(x)$
$(f(x) \pm g(x))' = f'(x) \pm g'(x)$
$\frac{d}{dx}(x^n) = nx^{n-1}$; $\frac{d}{dx}(c) = 0$
$\frac{d}{dx}(f(g(x))) = f'(g(x))g'(x)$
product: $(fg)' = f'g + fg'$; quotient: $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$
$\frac{d}{dx}\sin x = \cos x$; $\frac{d}{dx}\cos x = -\sin x$; $\frac{d}{dx}\tan x = \sec^2 x$
$\frac{d}{dx}\sec x = \sec x\tan x$; $\frac{d}{dx}\csc x = -\csc x\cot x$; $\frac{d}{dx}\cot x = -\csc^2 x$
$\frac{d}{dx}\sin^{-1}x = \frac{1}{\sqrt{1-x^2}}$; $\frac{d}{dx}\cos^{-1}x = \frac{-1}{\sqrt{1-x^2}}$; $\frac{d}{dx}\tan^{-1}x = \frac{1}{1+x^2}$
$\frac{d}{dx}\sec^{-1}x = \frac{1}{|x|\sqrt{x^2-1}}$; $\frac{d}{dx}\csc^{-1}x = \frac{-1}{|x|\sqrt{x^2-1}}$; $\cot^{-1}x = \frac{-1}{1+x^2}$
$\frac{d}{dx}(a^x) = a^x\ln(a)$; $\frac{d}{dx}e^x = e^x$
$\frac{d}{dx}(\ln(x)) = \frac{1}{x}$, $x > 0$; $\frac{d}{dx}(\ln|x|) = \frac{1}{x}$, $x \ne 0$
$\frac{d}{dx}(\log_a(x)) = \frac{1}{x\ln a}$, $x > 0$

Integrals:
$\int x^{-n}dx = \left(\frac{1}{-n+1}\right)x^{-n+1}$, $n \ne 1$
$\int \frac{1}{ax+b}dx = \frac{1}{a}\ln|ax+b|$; $\int x^{p/q}dx = \frac{q}{p+q}x^{\frac{p+q}{q}}$
$\int e^u du = e^u$; $\int a^u du = \frac{a^u}{\ln a}$; $\int \ln u\, du = u\ln u - u$
$\int ue^u du = (u-1)e^u$; $\int \frac{1}{u\ln u}du = \ln|\ln u|$

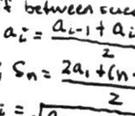
Whenever you divide or multiply by a negative number, you must flip the inequality sign!

Random point in unit circle:
X = distance to origin
Probability of point lying in a smaller circle C is $P(C) = \frac{\text{area } C}{\pi}$.
For $0 < x < 1$, event $\{X \le x\}$ is point lying in circle of radius x.
So $P(X \le x) = \frac{\pi x^2}{\pi} = x^2$, so cdf is $F_X(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \le x < 1 \\ 1, & 1 \le x \end{cases}$
pdf is $f_X(x) = \begin{cases} 2x, & 0 \le x < 1 \\ 0, & \text{elsewhere} \end{cases}$
$P(\frac{1}{4} < X \le \frac{1}{2}) = \int_{1/4}^{1/2} 2w\,dw = \frac{3}{16}$

Calls to switchboard
$f(x) = \begin{cases} \frac{e^{-x/4}}{4}, & 0 < x < \infty \\ 0, & \text{elsewhere} \end{cases}$



$f(x,y) = \begin{cases} 1, & 0 < x < 1,\ 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$
$Z = XY$; $F(z) = P(Z \le z)$
$= P(XY \le z) = P(Y \le \frac{z}{x})$ = volume under $f(x,y)$ over shaded area



$= \int_0^z\int_0^1 dy\,dx + \int_z^1\int_0^{z/x} dy\,dx$
$f(z) = \frac{d}{dz}F(z)$

$f(x) = \begin{cases} x^2/9, & 0 \le x < 3 \\ 0, & \text{elsewhere} \end{cases}$ $Y = X^3$
$F(Y) = P(Y \le y) = P(X^3 \le y)$
$= P(X \le \sqrt[3]{y}) = \int_0^{\sqrt[3]{y}} x^2/9\,dx$
$f(y) = \frac{d}{dy}F(Y)$ y support

$f(x) = \begin{cases} 1/3, & -1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$
$Y = X^2$; $F(Y) = P(Y \le y) = P(X^2 \le y)$
Two cases: ① $0 \le y < 1$; ② $1 \le y < 4$
① $P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y})$
$= \int_{-1}^{\sqrt{y}} 1/3\,dx - \int_{-1}^{-\sqrt{y}} 1/3\,dx = \frac{2\sqrt{y}}{3}$
② $P(X^2 \le y) = P(X \le \sqrt{y}) = \int_{-1}^{\sqrt{y}} 1/3\,dx$
$= \frac{1+\sqrt{y}}{3}$
$F(y) = \begin{cases} 0, & -\infty < y < 0 \\ 2\sqrt{y}/3, & 0 < y < 1 \\ (1+\sqrt{y})/3, & 1 < y < 4 \\ 1, & y \ge 4 \end{cases}$
$f(y) = \frac{d}{dy}F(y)$
$F(Y) = \begin{cases} \frac{1}{3\sqrt{y}}, & 0 < y < 1 \\ \frac{1}{6\sqrt{y}}, & 1 < y < 4 \end{cases}$

$p_{X_1X_2}(x_1,x_2) = \frac{\mu_1^{x_1}\mu_2^{x_2}e^{-\mu_1}e^{-\mu_2}}{x_1!\,x_2!}$ pmf, x_1 and x_2 nonneg. integers
We want pmf of $Y_1 = X_1 + X_2$. We need to define a Y_2, let's choose $Y_2 = X_2$.
Inverse functions are $x_1 = y_1 - y_2$ and $x_2 = y_2$. Thus, $p_{Y_1Y_2}(y_1,y_2) = \frac{\mu_1^{y_1-y_2}\mu_2^{y_2}e^{-\mu_1-\mu_2}}{(y_1-y_2)!\,y_2!}$
*Note: $y_2 = 0, 1, \dots, y_1$; $y_1 = 0, 1, 2, \dots$

Random point on unit square. We want not X or Y, but $Z = X+Y$. $f_{XY}(x,y) = \{1$, on $\}$; 0, elsewhere
$F_Z(z) = P(X+Y \le z) = \begin{cases} 0, & z < 0 \\ \int_0^z\int_0^{z-x} dy\,dx, & 0 < z < 1 \\ 1 - \int_{z-1}^1\int_{z-x}^1 dy\,dx, & 1 \le z < 2 \\ 1, & 2 \le z \end{cases}$