MATH 181A Review

Kevin Jacob

Winter 2025

Table Of Contents

1 Summary of Models

1.1 Discrete Models

- **Geometric Distribution** $X \sim Geom(p)$
  $X \in \{1, 2, 3, \dots\}$
  $P_X(k) = p(1-p)^{k-1}$
  $\mathbb{E}[X] = \frac{1}{p}$
  $Var(X) = \frac{1-p}{p^2}$

- **Binomial Distribution** $X \sim Binom(n, p)$
  $X \in \{0, 1, 2, \dots, n\}$
  $P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$
  $\mathbb{E}[X] = np$
  $Var(X) = np(1-p)$

- **Poisson Distribution** $X \sim Poisson(\lambda)$
  $X \in \{0, 1, 2, \dots\}$
  $P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$
  $\mathbb{E}[X] = \lambda$
  $Var(X) = \lambda$

- **Negative Binomial Distribution** $X \sim NegBinom(r, p)$
  $X \in \{r, r+1, r+2, \dots\}$
  $P_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$
  $\mathbb{E}[X] = \frac{r}{p}$
  $Var(X) = \frac{r(1-p)}{p^2}$
  Note Same as Geometric when $r = 1$

1.2 Continous Models

- **Uniform Distribution** $X \sim Unif(a, b)$
  $f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$
  $\mathbb{E}[X] = \frac{a+b}{2}$
  $Var(x) = \frac{(b-a)^2}{12}$

- **Exponential Distribution** $X \sim Exp(\lambda)$
  $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
  $\mathbb{E}[X] = \frac{1}{\lambda}$
  $Var(X) = \frac{1}{\lambda^2}$

- **Normal Distribution** $X \sim N(\mu, \sigma^2)$
  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
  $\mathbb{E}[X] = \mu$
  $Var(X) = \sigma^2$

- **Gamma Distribution** $X \sim Gamma(r, \lambda)$
  $f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, x > 0, r > 0, \lambda > 0$
  $\mathbb{E}[X] = \frac{r}{\lambda}$
  $Var(X) = \frac{r}{\lambda^2}$
  Note Same as Exponential when $r = 1$
  Gamma Function
  $\Gamma(r) = \int_0^{\infty} y^{r-1} e^{-y} dy$
  If $r \in \mathbb{N}$ then $\Gamma(r) = (r-1)!$

2 Method of Moments Estimators/Estimates (MME)

2.1 Identically Distributed and Independent Random Variables

Identically Distributed
$\mathbb{E}[X_1] = \mathbb{E}[X_2] = \cdots = \mathbb{E}[X_n] = \mathbb{E}[X]$
$Var(X_1) = Var(X_2) = \cdots = Var(X_n) = Var(X)$
Independence
$\mathbb{E}[X_i + X_j] = \mathbb{E}[X_i] + \mathbb{E}[X_j]$
$Var(X_i + X_j) = Var(X_i) + Var(X_j)$

2.2 Moments

	First Moment	Second Moment	...	ith moment
Theoretical Moments:	$\mathbb{E}[X]$	$\mathbb{E}[X^2]$	...	$\mathbb{E}[X^i]$
example: $X \sim N(\mu, \sigma^2)$	$\mathbb{E}[X] = \mu$	$\mathbb{E}[X^2] = \sigma^2 + \mu^2$		
Sample Moments:	$\frac{1}{n}\sum_{j=1}^{n} x_j$	$\frac{1}{n}\sum_{j=1}^{n} X_j^2$	...	$\frac{1}{n}\sum_{j=1}^{n} x_j^i$

Note: $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
For MME we set theoretical moment equal to sample moment

2.3 Examples

1. Let $X_1, \ldots, X_n$ be iid based on $X \sim Exp(\lambda)$. Find MME for λ
   $\mathbb{E}[X] = \frac{1}{\lambda}$
   $\frac{1}{\lambda} = \bar{X}$
   Thus, $\hat{\lambda} = \frac{1}{\bar{X}}$, assuming $\bar{X} \neq 0$

2. Let $y_1, \ldots, y_n$ ne a random sample from the density $f_Y(y;\theta) = \frac{2y}{\theta^2}, 0 \leq y \leq \theta$. Find MME for θ
   $\mathbb{E}[Y] = \int_o^\theta y \cdot \frac{2y}{\theta^2} dy = \frac{2}{3}\theta$
   Equating moments gives $\bar{y} = \frac{2}{3}\hat{\theta}$. Thus, $\hat{\theta} = \frac{3}{2}\bar{y}$

3. Find MME for μ and σ^2 for $X \sim N(\mu, \sigma^2)$ assuming both are unknown.
   Equating first moments: $\hat{\mu} = \bar{X}$
   Equating second moments: $\hat{\sigma^2} + \hat{\mu}^2 = \frac{1}{n}\sum X_i^2 = \bar{X^2}$
   Since $\hat{\mu} = \bar{X}$, we get $\hat{\mu^2} = \bar{X^2} - \bar{X}^2$

3 Maximum Likelihood Estimators/Estimates (MLE)

3.1 Likelihood Function

A pmf or pdf where x is viewed as given and the parameter(s) is viewed as the unknown
Ex. We flip a coin 10 times and get heads 6 times
$L(p) = f(x = 6, p) = \binom{10}{6} p^6 (1-p)^4$
$L'(p) = \binom{10}{6}[p^6 \cdot 4(1-p)^3(-1) + (1-p)^4 \cdot 6p^5]$
$0 = -4p^6(1-p)^3 + 6p^5(1-p)^4 \to 10p = 6 \to \hat{p}_{MLE} = 0.6$
Make sure to do second derivative test to show it is a max

3.2 Log Likelihood Function

Take ln of $L(p)$ to get $\ell(p)$
We can utilize log properties such as $\ln(a \cdot b) = \ln(a) + \ln(b)$ and $\ln(\frac{a}{b}) = \ln(a) - \ln(b)$
Ex. Coin flipes from above $L(p) = f(x = 6, p) = \binom{10}{6} p^6 (1-p)^4$
$\ell(p) = \ln(L(p)) = \ln(\binom{10}{6}) +^6 + \ln(1-p)^4$
$\ell'(p) = 0 + \frac{6p^5}{p^6} + \frac{4(1-p)^3(-1)}{(1-p)^4} = \frac{6}{p} - \frac{4}{1-p}$
$0 = \frac{6}{\hat{p}} - \frac{4}{1-\hat{p}} \to \hat{p} = 0.6$

3.3 Multiple Pieces of Data

We have iid data $(x_1, \ldots, x_n)$ from a RV $X \sim f_X(s;\theta)$
Here, $L(\theta) = f_{(X_1,\ldots,X_n)}(x_1 \text{ and } x_2 \text{ and } \ldots \text{and } x_n) = f_{X_1}(x_1;\theta) \cdot f_{X_2}(x_2;\theta) \ldots f_{X_n}(x_n;\theta)$
So, $L(\theta) = \prod_{i=1}^n f(x_i;\theta)$
Thus, $\ell(\theta) = \ln \prod_{i=1}^n f(x_i;\theta) = \prod_{i=1}^n (x_i;\theta)$
Ex. $X \sim Poisson(\lambda)$ and you have data $x_1 \ldots, x_n$
$L(\lambda) = \prod_{i=1}^n f(x_i;\lambda) = \prod_{i=1}^n \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$
$\ell(\lambda) = \sum_{i=1}^n (x_i;\lambda) = \sum_{i=1}^n \ln(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}) = \sum_{i=1}^n [^{-\lambda} + \ln \lambda^{x_i} - \ln(x_i!)] = \sum_{i=1}^n [-\lambda + x_i \ln \lambda - \ln(x_i!)]$

3.4 Indicator Functions

Let $I_{(a,b)} = I_{(a,b)}(x) = \begin{cases} 1 & x \in (a,b) \\ 0 & \text{else} \end{cases}$
Ex. $X \sim Unif(0,\theta)$ where $f_X(x;\theta) = \frac{1}{\theta}$ where $0 \leq x \leq \theta$
$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} I_{[x_i,\infty)} = \frac{1}{\theta^n} \prod I_{[x_i,\infty)} = \frac{1}{\theta^n} \cdot I_{[\max x_i,\infty)}(\theta)$
because of the $\frac{1}{\theta^n}$ term, $L(\theta)$ is maximal when θ is as small as possible. The indicator function forces $\max x_i \leq \theta < \infty$ (unless we want $L = 0$), so $\hat{\theta}_{MLE} = \max x_i$.

3.5 Multiple Parameters

$L(\theta_1, \theta_2) =$ some expression
$\ell(\theta_1, \theta_2) = \ln(\text{expression})$
Set $\frac{\partial \ell}{\partial \theta_1}$ and $\frac{\partial \ell}{\partial \theta_2}$ to 0.
Ex. Find the MLEs for r and λ in the Gamma distribution using the sample $x_1, \ldots, x_n$
$X \sim Gamma(r,\lambda) \qquad f_X(x;r,\lambda) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$ for $x > 0$
$L(r,\lambda) = \prod_{i=1}^n \frac{\lambda^r}{\Gamma(r)} x_i^{r-1} e^{-\lambda x_i}$
$\ell(r,\lambda) = \sum_{i=1}^n [r \ln \lambda \ln \Gamma(r) + (r-1)\ln x_i - \lambda x_i] = nr \ln \lambda - n \ln \Gamma(r) + (r-1)\sum \ln x_i - \lambda \sum x_i$
$\frac{\partial \ell}{\partial \lambda} = \frac{nr}{\lambda} + 0 + 0 - \sum x_i \to 0 = \frac{n\hat{r}}{\hat{\lambda}} - \sum x_i$ so $\hat{\lambda} = \frac{n\hat{r}}{\sum x_i} = \frac{\hat{r}}{\frac{\sum x_i}{n}} = \frac{\hat{r}}{(\bar{x})}$
$\frac{\partial \ell}{\partial r} = n \ln \lambda - n\frac{\Gamma'(r)}{\Gamma(r)} + \sum \ln x_i - 0 \to 0 = n \ln \hat{\lambda} - n \frac{\Gamma(\hat{r})}{\Gamma(\hat{r}) + \sum \ln x_i}$
Substituting $\hat{\lambda} = \frac{\hat{r}}{(\bar{x})}$ gives $0 = n \ln(\frac{\hat{r}}{(\bar{x})}) - n\frac{\Gamma'(\hat{r})}{\Gamma(\hat{r})} + \sum \ln x_i$ (no closed form solution)

4 Confidence Intervals

4.1 Empirical Rule

68% of the data lies within one standard deviation of the mean
95% of the data lies within 1.96 standard deviations of the mean
99.7% of the data lies within 3 standard deviations of the mean

4.2 Normal Distribution Mean

Make a confidence interval with n samples and σ^2 is known
$L = \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
$U = \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
$1 - \alpha$ *is the confidence level*
Statements about confidence interval:
incorrect: $\mu \in$ our CI
incorrect: There is a 95% chance that $\mu \in$ our CI
correct: 95% of the random CIs contain μ
correct: **If** $\mu \in$ our CI then ____ has an average between L and U

4.3 Proportions

Make a confidence interval with n samples for p
$L = \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
$U = \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

4.4 Working Backwards

$z_{\alpha/2} = \frac{d}{\sqrt{\frac{p(1-p)}{n}}} \to n = \frac{z_{\alpha/2}^2 \cdot p(1-p)}{d^2}$

- If we have no knowledge of p we set $p = \frac{1}{2}$ since $p(1-p)$ is maximal when $p = \frac{1}{2}$
- Make sure to always round up when finding n

5 Order Stats

Let $x_1, x_2, \ldots, x_n$ be iid from a CRV X
$x_{(i)}$ is the ith order statistic of the sample where $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$
PDF of the ith order statistic
$f_{X_{(i)}} = \frac{n!}{(i-1)!(n-i)!}F(x)^{i-1}f(x)[1-F(x)]^{n-i}$
Ex. $X \sim Unif(1,3)$ find pdf of $x_{(4)}$ based on iid data $x_1, \ldots, x_5$
$n = 5 \qquad i = 4$
$f_{x_{(4)}}(x) = \frac{5!}{(4-1)!(5-4)!}F_X(x)^{4-1}f_X(x)[1-F_X(x)]^{5-4}$
$= 20F_X(x)^3 f_X(x)[1-F_X(x)]$
$= 20(\frac{x-1}{2})^3(\frac{1}{2})[1 - \frac{x-1}{2}]$
$= 20\frac{(x-1)^3}{8}(\frac{1}{2})(\frac{3-x}{2})$
$= \frac{5}{8}(x-1)^3(3-x) \qquad 1 \leq x \leq 3$

6 Bias

Definition: The *bias* B of an estimator,$\hat{\theta}$, is $B = \mathbb{E}[\hat{\theta}] - \theta$
Fixing Bias: If $\mathbb{E}[\hat{\theta}] = c\theta$ or $\mathbb{E}[\hat{\theta}] = c + d$ we can just use $\frac{\hat{\theta}}{c}$ of $\hat{\theta} - d$ to compensate and create an unbiased estimator
Asymptotically Unbiased: Id $\hat{\theta}_n$ is an estimator for a sample size of n, We say $\hat{\theta}$ is *asymptotically unbiased* iff $\lim_{n\to\infty} \mathbb{E}[\hat{\theta}_n] = \theta$, or equivalently $\lim_{n\to\infty} B_n = 0$

Ex1. Suppose X and RV is modeled by $f(x;\theta) = \frac{3x^2}{\theta^3}$ where $0 \leq x \leq \theta$ and $\theta > 0$. For a sample $X_1, \ldots, X_n$ you can show that $\hat{\theta}_{MME} = \frac{4}{3}\bar{X}$. Show that $\hat{\theta}_{MME}$ is unbiased.
$\mathbb{E}[X] = \int_0^\theta x \cdot \frac{3x^2}{\theta^3}dx = \frac{3}{\theta^3}(\frac{1}{4}x^4)|_0^\theta = \frac{3}{4}\theta$
$\mathbb{E}[\hat{\theta}_{MME}] = \mathbb{E}[\frac{4}{3}\bar{X}] = \frac{4}{3}\mathbb{E}[\bar{X}] = \frac{4}{3}\mathbb{E}[X] = \frac{4}{3} \cdot \frac{3}{4}\theta = \theta$

Ex2. Show that MLE$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ is biased for σ^2 in $X \sim \mathcal{N}(\mu, \sigma^2)$
$\mathbb{E}[\hat{\sigma^2}] = \frac{1}{n}\mathbb{E}[\sum(X_i - \bar{X})^2] = \frac{1}{n}\mathbb{E}[\sum(X_i^2 - 2X_i\bar{X} + \bar{X}^2)] = \frac{1}{n}\mathbb{E}[\sum X_i^2 - 2\bar{X}\sum X_i + \sum \bar{X}^2]$
$= \frac{1}{n}\mathbb{E}[\sum X_i^2 - 2\bar{X} \cdot n\bar{X} + \sum \bar{X}^2] = \frac{1}{n}\mathbb{E}[\sum X_i^2 - 2n\bar{X}^2 + n\bar{X}^2] = \frac{1}{n}(\sum \mathbb{E}[X_i^2] - \mathbb{E}[n\bar{X}^2])$
$= \frac{1}{n}\sum \mathbb{E}[X^2] - \mathbb{E}[\bar{X}^2] = \frac{1}{n}n\mathbb{E}[X^2] - \mathbb{E}[\bar{X}^2] = \mathbb{E}[X^2] - \mathbb{E}[\bar{X}^2] = Var(X) + \mathbb{E}[X]^2 - (Var(\bar{X}) + \mathbb{E}[\bar{X}]^2)$
$= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2$

7 Efficiency

For two *unbiased estimators*, $\hat{\theta}_1$ is **more efficient** than $\hat{\theta}_2$ if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$.
The **relative efficiency** of $\hat{\theta}_1$ to $\hat{\theta}_2$ is $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}$
Ex. Let $X \sim Unif(0, \theta)$. We can show that $\hat{\theta}_1 = 2\bar{X}$ and $\hat{\theta}_2 = \frac{n+1}{n} X_{\max}$ are unbiased estimators of θ. Find the relative efficiency of $\hat{\theta}_2$ to $\hat{\theta}_1$.
1. Find $Var(\hat{\theta}_1)$
$Var(\hat{\theta}_1) = Var(2\bar{X}) = 4Var(\bar{X}) = \frac{4Var(X)}{n} = \frac{4 \cdot \frac{(\theta-0)^2}{12}}{n} = \frac{\theta^2}{3n}$
2. Find $Var(\hat{\theta}_2)$
$f_{x_i} = \frac{n!}{(i-1)!(n-i)!} = F_X(x)^{i-1} f_X(x)(1 - F_X(x))^{n-i}$
$F_X(x) = \int_0^\infty \frac{1}{\theta} dt = \frac{t}{\theta}\Big|_0^x = \frac{x}{\theta}$
$f_{X_{\max}}(x) = \frac{n!}{(n-1)!} (\frac{x}{\theta})^{n-1} \frac{1}{\theta} = \frac{nx^{n-1}}{\theta^n}$
$\mathbb{E}[X_{\max}^2] = \int_0^\theta x^2 \cdot \frac{nx^{n-1}}{\theta^n} dx = \frac{n\theta^2}{n+2}$
$Var(\hat{\theta}_2) = Var(\frac{n+1}{n} X_{\max}) = \mathbb{E}[(\frac{n+1}{n})^2 X_{\max}^2] - (\mathbb{E}[\frac{n+1}{n} X_{\max}])^2 = \frac{(n+1)^2}{n^2} \frac{n\theta^2}{n+2} - \theta^2 = \frac{\theta^2}{n(n+2)}$
$\text{eff}(\hat{\theta}_2, \hat{\theta}_1) = \frac{\frac{\theta^2}{3n}}{\frac{\theta^2}{n(n+2)}} = \frac{n+2}{3}$

8 Mean Squared Error

The **mean squared error** of an estimator $\hat{\theta}$ for a parameter θ is $MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$.
The **relative efficiency** is $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}$
$MSE(\hat{\theta}) = Var(\hat{\theta}) + (\text{Bias}(\hat{\theta}))^2$
Ex. Let $Y \sim Binom(n, p)$. Use relative efficiency to decide which is better. $\hat{p}_1 = \frac{Y}{n}$ and $\hat{p}_2 = \frac{Y+1}{n+1}$
1. Fine $MSE(\hat{p}_1)$
$MSE(\hat{p}_1) = Var(\frac{Y}{n}) + (\mathbb{E}[\frac{Y}{n} - p])^2 = \frac{1}{n^2} \cdot np(1-p) + (\frac{1}{n} \cdot np - p)^2 = \frac{p(1-p)}{n}$
2. Find $MSE(\hat{p}_2)$
$MSE(\hat{p}_2) = \frac{1}{(n+2)^2} \cdot np(1-p) + (\frac{1}{n+2} \cdot (np+1) - p)^2$
When $n = 4$, $\text{eff}(\hat{p}_2, \hat{p}_1) > 1$ so $\hat{p}_2$ is more efficient that $\hat{p}_1$ despite being a biased estimator

9 Fisher Information and Cramer-Rao Lower Bound

9.1 Fisher Information

Let X be a RV modeled by a smooth density function $f_X(x;\theta)$. We define *Fisher Information* as:
$I(\theta) = \mathbb{E}[(\frac{\partial \ln f(X;\theta)}{\partial \theta})^2] = -\mathbb{E}[\frac{\partial^2 \ln f(X;\theta)}{\partial \theta^2}]$
The Fisher Information, $I(\theta)$ gives a numerical sense for how much information we can squeeze out of a single data value, X. If the datum X will be very useful, I will be big; when the datum is not so helpful, I will be small.
Ex. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ unknown. Find the FI for μ.
$\ln f(X; \mu, \sigma^2) = -\frac{1}{2} \ln(2\pi) - \ln \sigma - \frac{(X-\mu)^2}{2\sigma^2}$
$\frac{\partial \ln f}{\partial \mu} = 0 - \frac{1}{2\sigma^2} \cdot 2(X - \mu)(-1) = \frac{X-\mu}{\sigma^2}$
$I(\mu) = \mathbb{E}_X[(\frac{X-\mu}{\sigma^2})^2] = \frac{1}{\sigma^4} \mathbb{E}[(X-\mu)^2] = \frac{1}{\sigma^4} \sigma^2 = \frac{1}{\sigma^2}$
$\frac{\partial^2 \ln f}{\partial \mu^2} = \frac{-1}{\sigma^2}$
$I(\mu) = -\mathbb{E}[\frac{\partial^2 \ln f}{\partial \mu^2}] = -\mathbb{E}[\frac{-1}{\sigma^2}] = \frac{1}{\sigma^2}$

9.2 Cramer-Rao Lower Bound

$Var(\hat{\theta}) \geq [nI(\theta)]^{-1}$
An unbiased estimator is said to be **efficient** iff its variance equals the CRLB
Ex. For $X \sim \mathcal{N}(0,1)$ we say that $I(\mu) = \frac{1}{\sigma^2}$. Find the Cramer-Rao Lower Bound
CRLB $= [nI(\mu)]^{-1} = \frac{\sigma^2}{n}$

10 Consistency

10.1 Regular Consistency

A sequence of estimators, $\hat{\theta_n}$, is **consistent for a value** θ iff for all $\epsilon > 0$:

$\lim_{n\to\infty} \mathbb{P}(|\hat{\theta_n} - \theta| < \epsilon) = 1$ or equivalently $lim_{n\to\infty}\mathbb{P}(|\hat{\theta_n} - \theta| > \epsilon) = 0$

Ex. Determine if $\hat{\theta_n} = X_{\min}$ is consistent for θ in the shifted exponential model
$X \sim f(x;\theta) = e^{-(x-\theta)}, x \geq \theta$
$F(x;\theta) = \int_\theta^x e^{-t+\theta} dt = -e^{-t+\theta}\big|_\theta^x = 1 - e^{-x+\theta}$
$f_{X_{\min}} = \frac{n!}{0!(n-1)!} f_X(x) \cdot (1 - F_X(x))^{n-1} = ne^{-x+\theta}(e^{-x+\theta})^{n-1} = ne^{-nx+n\theta}$
$\mathbb{P}(|\hat{\theta_n} - \theta| < \epsilon) = \mathbb{P}(\theta - \epsilon < X_{\min} < \theta + \epsilon) = \int_\theta^{\theta+\epsilon} ne^{-nx+n\theta} = -e^{-nx+n\theta}\big|_\theta^{\theta+\epsilon} = 1 - e^{-n\epsilon} \to 1$ as $n \to \infty$

10.2 MSE Consistency

A sequence of estimators $\hat{\theta_n}$ is (mean) squared error consistent iff

$\lim_{n\to\infty} \mathbb{E}[(\hat{\theta_n} - \theta)^2] = 0$, or equivalently $\lim_{n\to\infty} MSE(\hat{\theta_n}, \theta) = 0$

If $\hat{\theta_n}$ is MSE consistent for θ, the $\hat{\theta_n}$ is consistent of θ
Ex. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with σ^2 know. Show that $\bar{X}$ is consistent for μ by showing $\bar{X}$ is MSE consistent for μ.
$MSE(\bar{X}) = Var(\bar{X}) + \text{Bias}(\bar{X})^2 = \frac{Var(X)}{n} + (\mathbb{E}[\bar{X}] - \mu)^2 = \frac{\sigma^2}{n} + (\mathbb{E}[\bar{X}] - \mu)^2 = \frac{\sigma^2}{n} + 0^2 \to 0$ as $n \to \infty$
Since $\bar{X}$ is MSE-consistent for μ, it is also consistent for μ

11 Confidence Intervals MLE

$\hat{\theta_{\text{MLE}}} \approx \mathcal{N}(\theta_0, \frac{1}{nI(\theta_0)})$
$\frac{\hat{\theta_{\text{MLE}}} - \theta_0}{1/\sqrt{nI(\theta_0)}} \approx \mathcal{N}(0,1) \qquad 1 - \alpha \approx \mathbb{P}(-z_{\alpha/2} < \frac{\hat{\theta_{\text{MLE}}} - \theta_0}{1/\sqrt{nI(\theta_0)}} < z_{\alpha/s})$
$1 - \alpha \approx \mathbb{P}(\hat{\theta_{\text{MLE}}} - z_{\alpha/2}\frac{1}{\sqrt{nI(\theta_0)}} < \theta_0 < \hat{\theta_{\text{MLE}}} + z_{\alpha/2}\frac{1}{\sqrt{nI(\theta_0)}})$
A $(1-\alpha)100\%$ approximate CI for θ_0 is $(\hat{\theta_{\text{MLE}}} - z_{\alpha/2}\frac{1}{\sqrt{nI(\theta_0)}}, \hat{\theta_{\text{MLE}}} + z_{\alpha/2}\frac{1}{\sqrt{nI(\theta_0)}})$
Ex. Suppose $X \sim Poisson(\lambda)$ and we want an approximate 95% confidence interval based on 9 samples where $\bar{X} = \frac{16}{9}$
$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \ln f(X;\lambda) = -\lambda + X\ln\lambda - \ln(X!)$
$\frac{\partial \ln f}{\partial \lambda} = -1 + \frac{X}{\lambda} \qquad \frac{\partial^2 \ln f}{\partial \lambda^2} = -\frac{X}{\lambda^2} \qquad I(\lambda) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$
$\hat{\lambda_{\text{MLE}}} = \bar{X}, I(\hat{\lambda}) = \frac{1}{\bar{X}}$
MLE CI is $\bar{X} \pm 1.96\sqrt{\frac{\bar{X}}{n}}$
$\frac{16}{9} \pm 1.96\sqrt{\frac{16}{9\cdot 9}} \approx (0.907, 2.649)$

12 Hypothesis Test

12.1 Structure

1. Define parameters and set up hypothesis (H_0 and H_1)
2. Assume H_0 find the distribution of the test statistic
3. Describe how strange your data/stat are on this distribution
4. Make choice between H_0 and H_1, and express carefully (seems to be)

12.2 Mean

$H_0 : \mu = 284 \qquad H_1 : \mu < 284$
$\sigma^2 = 35^2 \qquad n = 100$
The distribution of the test stat is $\bar{X} \sim \mathcal{N}(284, \frac{35^2}{100})$
In the sample $\bar{X} = 274$
P-Value $= \mathbb{P}(\bar{X} \leq 274) = \mathbb{P}(z \leq \frac{274-284}{35/10}) \approx 0.002 \qquad z \sim \mathcal{N}(0,1)$
Since $0.002 < \alpha = 0.05$ we reject H_0 in favor of H_1. The average seems lower.
For a 2-sided test $\neq$ the p-value is the area under both tails based on symmetry

12.3 Proportion

$X \sim Binom(n,p)$
Option 1: Use x on $X \sim Binom(n,p)$
Option 2: $x \to \frac{x}{n} = \hat{p} \to$ use $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$
We can use normal approximation if oneof the following are true

- $0 < np - 2\sqrt{np(1-p)} < np + 3\sqrt{np(1-p)} < n$ (Larsen & Marx)
- $np, n(1-p) \geq 10$

$H_0 : p = \frac{1}{4} \qquad H_1 : p < \frac{1}{4}$
$\hat{p} = \frac{60}{747} \qquad n = 747 \qquad z \sim \mathcal{N}(0,1)$
P-Value $= \mathbb{P}(z \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}) = \mathbb{P}(z \leq \frac{\frac{60}{747} - \frac{1}{4}}{\sqrt{\frac{\frac{1}{4}(\frac{3}{4})}{747}}}) = \mathbb{P}(z \leq -10.710) \approx 4.568 \times 10^{-27}$
Since P-Value $< \alpha = 0.05$ we reject H_0 in favor of H_1

12.4 Duality

12.4.1 Two Sided

$X \sim \mathcal{N}(\mu, \sigma^2)$
μ_0 is reasonable based on a HT $\iff$ We keep $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$
$\iff -z_{\alpha/2} <$ test stat $< z_{\alpha/2} \iff -z_{\alpha/2} < \frac{\bar{X}-\mu+0}{\sigma/\sqrt{n}} < z_{\alpha/2}$
$\iff \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \iff \mu_0 \in \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
$\iff \mu_0 \in$ our CI $\iff \mu_0$ is reasonable based on a CI

12.4.2 One Sided

$X \sim \mathcal{N}(\mu, \sigma^2)$
We keep $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0 \iff$ test stat $< z_\alpha \iff \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < z_\alpha$
$\iff \bar{X} - \mu_0 < z_\alpha \frac{\sigma}{\sqrt{n}} \iff \bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}} < \mu_0 \iff \mu_0 \in$ CI given by $(\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty)$

12.5 Error

Type 1 Error: $\mathbb{P}(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$
Type 2 Error: $\mathbb{P}(\text{Keep } H_0 | H_1 \text{ is true}) = \beta$
Ex. $X \sim \mathcal{N}(\mu, 1.2^2) \qquad H_0 : \mu = 5 \qquad H_1 : \mu > 5 \qquad n = 40 \qquad \alpha = 0.06 \qquad \mu$ is actually 5.4
$\bar{X} \sim \mathcal{N}(5, \frac{1.2^2}{40})$
Type One Error $= \alpha = 0.06$
$0.06 = \mathbb{P}(z > \frac{C-5}{1.2/\sqrt{40}}) \to C \approx 5.292$
Type Two Error $= \mathbb{P}(\text{Keep } H_0 | H_1 \text{ is true}) = \mathbb{P}(z < \frac{C-5.4}{1.2/\sqrt{40}}) \approx 29\%$

12.6 Power

Power allows us to find the probability that the experiment will lead us to H_1 rather than H_0 if H_1 is actually true
Power $= \mathbb{P}(\text{reject } H_0 | H_1 \text{ is true}) = 1 - \mathbb{P}(\text{Keep } H_0|H_1 \text{ is true}) = 1 - \beta$
Larger n will make larger power and smaller β and as α increases, β decreases
Ex. $X \sim \mathcal{N}(\mu, \frac{1385^2}{n})$ $\quad H_0 : \mu = 7473 \quad H_1 : \mu > 7473 \quad \alpha = 0.05 \quad \beta = 0.01 \quad \mu$ is actually 9656
Find n
Start by finding C based on the type 2 error
$0.1 = \mathbb{P}(\bar{X} \leq C) = \mathbb{P}(\frac{\bar{X}-9656}{1385/\sqrt{n}} \leq \frac{C-9656}{1385/\sqrt{n}}) = \mathbb{P}(z \leq \frac{C-9656}{1385/\sqrt{n}})$
$\Phi^{-1}(0.01) = \frac{X-9656}{1385/\sqrt{n}}$
$C = 9656 + \Phi^{-1}(0.01)(\frac{1385}{\sqrt{n}})$
Find C based on α
$0.05 = \mathbb{P}(\frac{\bar{X}-7473}{1385/\sqrt{n}} > \frac{C-7473}{1385/\sqrt{n}})$
$\frac{C-7473}{1385/\sqrt{n}} = \Phi^{-1}(0.95)$
$C = 7473 + \Phi^{-1}(0.95)\frac{1385}{\sqrt{n}}$
Solve for n which gives $n \approx 3.449$ and we round up to $n \geq 4$ since $n = 4$ is the smallest sample size to guarantee our requirements.

12.7 Other Estimators

We can utilize the distribution of another estimator to calculate the critical value and perform our hypothesis test
Ex. $h_0 : \theta = 3 \quad H_1 : \theta < 3 \quad n = 4 \quad \alpha = 0.06 \quad \theta$ is actually 2.5 $\quad$ Sample stat for test is X_{MAX}
$f(x;\theta) = \frac{2x}{\theta^2}, 0 \leq x \leq \theta$
$F_X(x) = \int_0^x \frac{2t}{\theta^2} dt = \frac{t^2}{\theta^2}\Big|_0^x = \frac{x^2}{\theta^2}$
$f_{X_{\text{MAX}}}(x;\theta) = \frac{n!}{(n-1)!} F_X(x)^{n-1} f_X(x) = 4(\frac{x^2}{\theta^2})^3 \frac{2x}{\theta^2} = \frac{8x^7}{\theta^8}$
$0.06 = \int_0^c f_{X_{\text{MAX}}}(x;\theta = 3)dx = \int_0^c \frac{8x^7}{3^8} dx = \frac{x^8}{3^8}\Big|_0^c = (\frac{c}{3})^8 \to c \approx 2.11$
Find type 2 error ($\theta = 2.5$)
$\beta = \int_c^{2.5} f_{X_{\text{MAX}}}(x;\theta = 2.5)dx = \int_{2.11}^{2.5} \frac{8x^7}{2.5^8} dx \approx 0.742$

13 chi-square, F, and t distribution

13.1 chi-square $\mathcal{X}^2$

$Z \sim \mathcal{N}(0, 1) \qquad Z_1, \ldots, Z_n$
$\mathcal{X}^2 \sim \sum_{i=1}^n Z_i^2$
The pdf for $\mathcal{X}_n^2$ is $f(u) = \frac{u^{n/2-1} \cdot e^{=u/2}}{2^{n/2} \cdot \Gamma(\frac{n}{2})}$ which is the pdf of $Gamma(r = \frac{n}{2}, \lambda = \frac{1}{2})$
Other chi-square facts

- $\mathcal{X}_n^2 + \mathcal{X}_m^2 = \mathcal{X}_{n+m}^2$
- $\mathbb{E}[\mathcal{X}_\backslash^\in] = n$
- $Var(\mathcal{X}_n^2) = 2n$

Ex.
$S^2 = \frac{1}{n-1} \sum (X_i - \overline{X})^2 \qquad$ Find the distribution of $\frac{(n-1)S^2}{\sigma^2}$
$\sum_{i=1}^n (\frac{X_i - \mu}{\sigma})^2 = \frac{1}{\sigma^2} \sum (X_i - \overline{X} + \overline{X} - \mu)^2 = \frac{1}{\sigma^2} \sum (X_i - \overline{X})^2 + \frac{2}{\sigma^2} \sum (X_i - \overline{X})(\overline{X} - \mu) + \frac{1}{\sigma^2} \sum (\overline{X} - \mu)^2$
The first term is $\frac{(n-1)S^2}{\sigma^2}$ abd the last term is $\frac{n(\overline{X}-\mu)^2}{\sigma^2} = (\frac{\overline{X}-\mu}{\sigma/\sqrt{n}})^2$
The middle term is $\frac{2(\overline{X}-\mu)}{\sigma^2} \sum (X_i - \overline{X}) = \frac{2(\overline{X}-\mu)}{\sigma^2} [n\overline{X} - n\overline{X}] = 0$
Since $\frac{X_i - \mu}{\sigma} \sim Z_i$, the left hand side distribution is $\mathcal{X}_n^2$
The last term on the right side has distribution $Z^2 = \mathcal{X}_1^2$
Subtracting shows that $\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}_{n-1}^2$

13.2 $F_{m,n}$ distribution

Let $U = \mathcal{X}_n^2$ and $V = \mathcal{X}_m^2$ be independent chi-square random variables.
$F_{m,n} = \frac{V/m}{U/n}$
The pdf of the $F_{m,n}$ distribution is $f_{F_{m,n}} = \frac{\Gamma(\frac{m+n}{2})m^{m/2}n^{n/2}w^{(m/2)-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})(n+mw)^{(m+n)/2}}, w \geq 0$

13.3 T distribution T_n

Let $Z = \mathcal{N}(0,1)$ and $U = \mathcal{X}_n^2$
$T_n = \frac{Z}{\sqrt{\frac{U}{n}}}$
$T_n^2 = \frac{Z^2}{U/n} = \frac{\mathcal{X}_1^2/1}{\mathcal{X}_n^2/n} = F_{1,n}$
Ex.
Show that $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ is a t distribution
$\frac{\overline{X}-\mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\frac{S}{\sigma}} = \frac{Z}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{Z}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{(n-1)}}} = \frac{Z}{\sqrt{\frac{\mathcal{X}_{n-1}^2}{n-1}}} = T_{n-1}$

14 Hypothesis Test pt.2

14.1 Mean with Unknown Variance

$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$
Confidence Interval
$1-\alpha = \mathbb{P}(-t_{\alpha/2,n-1} \leq T_{n-1} \leq t_{\alpha/2,n-1}) = \mathbb{P}(-t_{\alpha/2,n-1} \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2,n-1})$
$= \mathbb{P}(\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}})$

$\overline{X} \pm t_{\alpha/n,n-1}\frac{S}{\sqrt{n}}$ is a $(1-\alpha)\%$ CI for μ
Hypothesis Test
H_0:$\mu = \mu_0$ $\qquad$ H_1:$\mu > \mu_0$
First compute $t = \frac{\overline{x}-\mu_0}{s/\sqrt{n}}$

1. If $t > t_{a,n-1}$. reject H_0; else keep H_0
2. If P-value=$\mathbb{P}(t < T_{n-1}) < \alpha$ reject H_0; else keep H_0

Need $n \geq 30$ **For T Approximation**

14.1.1 Variance

$\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}_{n-1}^2$
Confidence Interval
$1-\alpha = \mathbb{P}(\mathcal{X}_{\alpha/2,n-1}^2 \leq \mathcal{X}_{n-1}^2 \leq \mathcal{X}_{1-\alpha/2,n-1}^2) = \mathbb{P}(\mathcal{X}_{\alpha/2,n-1}^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \mathcal{X}_{1-\alpha/2,n-1}^2)$
$= \mathbb{P}(\frac{(n-1)S^2}{\mathcal{X}_{1-\alpha/2,n-1}^2} \leq \sigma \leq \frac{(n-1)S^2}{\mathcal{X}_{\alpha/2,n-1}^2})$

$(\frac{(n-1)S^2}{\mathcal{X}_{1-\alpha/2,n-1}^2}, \frac{(n-1)S^2}{\mathcal{X}_{\alpha/2,n-1}^2})$ is a $(1-\alpha)\%$ CI for σ^2
Hypothesis Test
$\mathcal{X}^2 = \frac{(n-1)S^2}{\sigma_0^2}$ and H_0:$\sigma^2 = \sigma_0^2$

- H_1:$\sigma^2 < \sigma_0^2$ reject H_0 if $\mathcal{X}^2 < \mathcal{X}_{\alpha,n-1}^2$ of if $\mathbb{P}(\mathcal{X}_{n-1}^2 < \mathcal{X}^2) < \alpha$
- H_1:$\sigma^2 > \sigma_0^2$ reject H_0 if $\mathcal{X}^2 > \mathcal{X}_{1-\alpha,n-1}^2$ of if $\mathbb{P}(\mathcal{X}_{n-1}^2 > \mathcal{X}^2) < \alpha$
- H_1:$\sigma^2 \neq \sigma_0^2$, reject H_0 if $\mathcal{X}^2 > \mathcal{X}_{1-\alpha/2,n-1}^2$ of $\mathcal{X}^2 < \mathcal{X}_{\alpha/2,n-1}^2$ (can't use P-value approach since chi-square is asymmetric)

15 Critical Regions

15.1 Best Critical Region

Let $X \sim f(x;\theta)$ and suppose S is the set of possible values for the random sample $\mathbf{X} = (X_1, \ldots, X_n)$.
Let $C \subseteq S$.
C is a **best critical region of size** α for testing the simple $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ iff

- $P_0(\mathbf{X} \in C) = \alpha$
- For $A \subseteq S$ with $P_0(\mathbf{X} \in A) = \alpha$ we have $P_1(\mathbf{X} \in C) \geq P_1(\mathbf{X} \in A)$

15.2 Likelihood Ratio

$\frac{P_0(X=v)}{P_1(X=v)}$. If $LR > 1$ the data are more likely to come from H_0 and if $LR < 1$, H_1 has a higher likelihood.

15.3 Neyman-Pearson Lemma (NPL)

Fix n, let $X \sim f(x;\theta)$ and draw a sample $\mathbf{X} = (X_1, \ldots, X_n)$. The likelihood of $\mathbf{X}$ is $L(\theta;x) = \prod_{i=1}^n f(x_i;\theta)$. Let k be any fixed positive real number.

Set up $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$. Create a set C via the following rule:

$$\text{If } \frac{L(\theta_0;x)}{L(\theta_1;x)} < k, \text{ put } \mathbf{x} \text{ in C; else don't}$$

The NPL claims that C will be a BCR of size $\alpha = P_0(\mathbf{X} \in C)$
Ex:
Let $X \sim f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$ and draw a sample size of n to test $H_0 : \theta = 1$ vs. $H_1 : \theta = 3$. Show that every BCR will take the form $C = \{\mathbf{X} | c < \prod_{i=1}^n X_i\}$
$L(\theta;x) = \prod_{i=1}^n (\theta x_i^{\theta-1}) = \theta^n (\prod x_i)^{\theta-1}$
$\frac{L(\theta=1)}{L(\theta=3)} = \frac{1^n (\prod x_i)^0}{3^n (\prod x_i)^2} < k$
$\frac{1}{k3^n} < (\prod x_i)^2$, or $c = \sqrt{\frac{1}{k3^n}} < \prod x_i$

15.4 Uniformly Most Powerful Critical Region (UMPCR)

A critical region C is a uniformly most powerful critical region of size α for testing the simple hypothesis H_0 against the composite hypothesis H_1 iff C is a BCR of size α for testing H_0 against each simple hypothesis in H_1
Ex.
Let $X \sim Exp(\lambda)$ and draw a sample size 3 for testing $H_0 : \lambda = 2$ vs. $H_1 : \lambda > 2$. Show the BCR looks like $\{\mathbf{X} | \sum X_i < c\}$ for any simple-simple comparison
$L(\lambda;x) \prod_{i=1}^3 \lambda e^{-\lambda x_i} = \lambda^3 e^{-\lambda \sum_{i=1}^3 x_i}$
$\frac{L(\lambda=2)}{L(\lambda=a)} < k \longleftrightarrow \frac{2^3 e^{-2\sum x_i}}{a^3 e^{-a\sum x_i}} = c_1 e^{(a-2)\sum x_i} < k \longleftrightarrow (a-2)\sum x_i < \ln(k/c_1) = c_2$
Since $a - 2 > 0$ we get $\sum x_i < \frac{c_2}{a-2} = c$

15.5 Generalized Likelihood Ratio (GLR) Tests

Let $X \sim f(x;\theta_1, \ldots, \theta_r)$ and $x_1, \ldots, x_n$ the values of a random sample. Let $\daleth \subseteq \mathbb{R}^n$ ne the set of values for $\theta_1, \ldots, \theta_r$ allow by H_0, and $\Omega \subseteq \mathbb{R}^r$ be those allowed by either H_0 of H_1. The generalized likelihood ratio is defined as follows
$\lambda = \frac{max_{\theta \in \daleth} L(\theta_1,\ldots,\theta_r;x_1,\ldots,x_n)}{max_{\theta \in \Omega} L(\theta_1,\ldots,\theta_r;x_1,\ldots,x_n)}$
We can create a critical region
$C = \{\mathbf{X} | \text{GLR} = \frac{max_{\theta \in \daleth} L(\theta_1,\ldots,\theta_r;x_1,\ldots,x_n)}{max_{\theta \in \Omega} L(\theta_1,\ldots,\theta_r;x_1,\ldots,x_n)} < k\}$

16 Bayesian Statistics

16.1 Beta Distribution

$\theta \sim Beta(a, b)$ where $a, b > 0$. By definition,
$f_\theta(\theta, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ where $0 \leq \theta \leq 1$

16.2 Posterior

$g(\theta|x) = \frac{f_X(x|\theta)\cdot g(\theta)}{f_X(x)}$
Ex.
$g_{Beta(5,3)}$
16 tack flips $Binom(16, \theta) \sim f(x; \theta) = \binom{16}{x}\theta^x(1-\theta)^{16-x}$
14 flips landed upwards
$g(\theta|x = 14) = \frac{f(x=14|\theta\cdot)g(\theta)}{f(x=14)}\alpha f(x = 14|\theta) \cdot g(\theta)\alpha\binom{16}{14}\theta^{14}(1-\theta)^2 \cdot \frac{\Gamma(5+3)}{\Gamma(5)\Gamma(3)}\theta^{5-1}(1-\theta)^{3-1}$
$\alpha\theta^{18}(1-\theta)^4$ kernel of $Beta(19, 5)$, so $g(\theta|14) \sim Beta(19, 5)$
α is a proportionality symbol we can use to hide constants

16.3 Conjugate Priors

If you get the same distribution for the prior and the posterior, then there is a "conjugate prior". In the prvious example, *Beta* is a conjugate prior for the *Binomial* likelihood.
Ex.
Show that $Beta(a, b)$ is a conjugate prior to $Geom(p)$ and find how hyperparameters get updated in the posterior distribution for the data $x_1, \ldots, x_n$.
$X \sim Geom(p)$ $\qquad f(\mathbf{x}; p) = \prod_{i=1}^n p(1-p)^{x_i-1} = p^n(1-p)^{(\sum x_i)-n}$
$g(p|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}p^{a-1}(1-p)^{b-1}$
$g(p|\mathbf{x})\alpha f(\mathbf{x}; p) \cdot g(p)\alpha p^n(1-p)^{(\sum x_i)-n} \cdot p^{a-1}(1-p)^{b-1}$
$\alpha p^{n+a-1}(1-p)^{(\sum x_i)-n+b-1} \sim Beta(a + n, b - n + \sum x_i)$
Updating rule is $Beta(a, b) \to Beta(a + n, b - n + \sum_{i=1}^n x_i)$