**Student ID (NOT your name):**

## Final Examination: QUESTION BOOKLET

Prof. Will Fithian

Fall 2021

---

- Do *NOT* open this question booklet until you are told to do so.
- Write your Student ID number (**NOT** your name) at the top of this page.
- Write your solutions in this booklet.
- No electronic devices are allowed during the exam.
- Be neat! If we can't read it, we can't grade it.
- You can treat any results from lecture or homework as "known," and use them in your work without rederiving them, but do make clear what result you're using. You do not need to explicitly check regularity conditions for the theorems from class that required them.
- For a multi-part problem, you may treat the results of previous parts as given (if you don't prove the result for part (a), you can still use it to solve part (b)).
- I have starred some parts which I believe are the most difficult, and which I expect most students won't necessarily be able to solve in the time allotted. They are generally not worth more points than the less difficult parts, so don't waste too much time on them until you're happy with your answers to the latter.
- Be careful to justify your reasoning and answers. We are primarily interested in your understanding of concepts, so show us what you know.
- Good luck!

**1. Regression with correlated errors (25 points, 5 points / part).**

Some useful facts for this problem:

- For $\mu \in \mathbb{R}^n$ and positive definite $\Sigma \in \mathbb{R}^{n\times n}$, the density for $Z \sim N_n(\mu, \Sigma)$ is
$$p_{\mu,\Sigma}(z) = |2\pi\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(z-\mu)'\Sigma^{-1}(z-\mu)\right\},$$
where $|\cdot|$ is the determinant (note the exponent of $1/2$ is correct; it should not be $n/2$). The mean is $\mu$ and the variance is $\Sigma$.

- If $Z \sim N_n(\mu, \Sigma)$, and $A \in \mathbb{R}^{k\times n}$ and $b \in \mathbb{R}^k$ are fixed, then
$$AZ + b \sim N_k(A\mu + b, A\Sigma A').$$

Suppose that for $i = 1, \ldots, n$ we observe fixed covariates $x_i \in \mathbb{R}^d$ and random response $Y_i = x_i'\beta + \varepsilon_i$, for coefficient vector $\beta \in \mathbb{R}^d$ and $\varepsilon_i \in \mathbb{R}$. The errors are multivariate Gaussian with mean zero and positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. In terms of the full response vector $Y \in \mathbb{R}^n$ and design matrix $X \in \mathbb{R}^{n\times d}$ with $i$th row $x_i'$, we have
$$Y = X\beta + \varepsilon, \quad \text{with } \varepsilon \sim N_n(0, \Sigma).$$

Assume $n \geq d \geq 1$ and $X$ has full column rank. For parts (a) and (b), we will assume $\Sigma$ is known and we want to estimate $\beta$. For (c)-(e) we will assume $\Sigma$ is unknown.

(a) Show that $Y$ follows a full-rank exponential family model and identify its complete sufficient statistic.

**Solution:**
We can write the density for $Y$ as
$$\begin{aligned} p_\beta(y) &= |2\pi\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y-X\beta)'\Sigma^{-1}(y-X\beta)\right\} \\ &= \exp\left\{\beta'X'\Sigma^{-1}y - \frac{1}{2}\beta'X'\Sigma^{-1}X\beta\right\} \cdot |2\pi\Sigma|^{-1/2} \exp\{-y'\Sigma^{-1}y/2\} \\ &= e^{\beta'T(y)-A(\beta)}h(y), \end{aligned}$$
for $T(y) = X'\Sigma^{-1}y$, $h(y) = |2\pi\Sigma|^{-1/2}\exp\{-y'\Sigma^{-1}y/2\}$, and $A(\beta) = \beta'X'\Sigma^{-1}X\beta/2$. Because the natural parameter $\beta$ can range over all of $\mathbb{R}^d$, the exponential family is full-rank and $T(Y)$ is complete.

(b) Find the maximum likelihood estimator of $\beta$ and give its distribution.

**Solution:**
The MLE for an exponential family sets

$$T(Y) = X'\Sigma^{-1}Y = \mathbb{E}_{\hat{\beta}}T(Y) = X'\Sigma^{-1}X\hat{\beta} \iff \hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y.$$

Its distribution, using the formula given in the preamble, is

$$\hat{\beta} \sim N\left(\beta, (X'\Sigma^{-1}X)^{-1}\right).$$

(c) Now, for the remainder of the problem, suppose that $\Sigma$ is unknown so we have to estimate it. To facilitate this, we observe i.i.d. replicates $Y^{(k)}$ for $k = 1, \ldots, m$, with distribution

$$Y^{(k)} = X\beta + \varepsilon^{(k)}, \quad \text{with } \varepsilon^{(k)} \overset{\text{i.i.d.}}{\sim} N_n(0, \Sigma).$$

Note that $X$ and $\beta$ are the same for $k = 1, \ldots, m$ (they do not depend on $k$); only the errors change (and the responses change as a result). Define

$$\overline{Y} = \frac{1}{m}\sum_{k=1}^{m} Y^{(k)}, \quad \text{and } \widehat{\Sigma} = \frac{1}{m-1}\sum_{k=1}^{m}(Y^{(k)} - \overline{Y})(Y^{(k)} - \overline{Y})'.$$

Show that $\overline{Y}$ and $\widehat{\Sigma}$ are independent of each other.

**Solution:**
Consider the model with $Y^{(1)}, \ldots, Y^{(m)} \overset{\text{i.i.d.}}{\sim} N_n(\mu, \Sigma)$, with arbitrary $\mu \in \mathbb{R}^n$ and positive definite $\Sigma$. In the submodel where $\Sigma$ is known, $\Sigma^{-1}\overline{Y}$ is complete sufficient (applying part (a) with $X = I_n$) and $\widehat{\Sigma}$ is ancillary, so by Basu's theorem $\Sigma^{-1}\overline{Y}$, and therefore also $\overline{Y}$, is independent of $\widehat{\Sigma}$. The two statistics are therefore independent for any $\mu$ and $\Sigma$, so in particular they are independent if $\mu = X\beta$ for any $\Sigma$.

**Common mistake:** $\overline{Y}$ is not complete sufficient in the model with $\mu = X\beta$, for $d < n$.

(d) Show that $\widehat{\Sigma}$ is an unbiased estimator of $\Sigma$.

**Solution:**

Note that $Y^{(k)} - \overline{Y} = \varepsilon^{(k)} - \bar{\varepsilon}$. For every $i, j \in \{1, \dots, d\}$ we have

$$\begin{aligned}
\mathbb{E}\widehat{\Sigma}_{ij} &= \frac{1}{m-1}\mathbb{E}\left[\sum_{k=1}^{m}(\varepsilon_i^{(k)} - \bar{\varepsilon}_i)(\varepsilon_j^{(k)} - \bar{\varepsilon}_j)\right] \\
&= \frac{m}{m-1}\mathbb{E}\left[\left(\frac{m-1}{m}\varepsilon_i^{(1)} - \frac{1}{m}\sum_{k>1}\varepsilon_i^{(k)}\right)\left(\frac{m-1}{m}\varepsilon_j^{(1)} - \frac{1}{m}\sum_{k>1}\varepsilon_j^{(k)}\right)\right]] \\
&= \frac{m}{m-1}\cdot\left[\left(\frac{m-1}{m}\right)^2\mathbb{E}\left[\varepsilon_i^{(1)}\varepsilon_j^{(1)}\right] + \sum_{k>1}\frac{1}{m^2}\mathbb{E}\left[\varepsilon_i^{(k)}\varepsilon_j^{(k)}\right]\right] \\
&= \mathbb{E}[\varepsilon_i^{(1)}\varepsilon_j^{(1)}] = \Sigma_{ij}.
\end{aligned}$$

An alternative way to do it is to observe that

$$\sum_{k=1}^{m}(\varepsilon^{(k)} - \bar{\varepsilon})(\varepsilon^{(k)} - \bar{\varepsilon})' = \left(\sum_{k=1}^{m}\varepsilon^{(k)}\varepsilon^{(k)\prime}\right) - m\bar{\varepsilon}\bar{\varepsilon}'.$$

Then, because $\mathbb{E}\left[\bar{\varepsilon}\bar{\varepsilon}'\right] = \text{Var}(\bar{\varepsilon}) = m^{-1}\Sigma$, we have

$$\begin{aligned}
\mathbb{E}\widehat{\Sigma}_{ij} &= \frac{1}{m-1}\mathbb{E}\left[\sum_{k=1}^{m}(\varepsilon_i^{(k)} - \bar{\varepsilon}_i)(\varepsilon_j^{(k)} - \bar{\varepsilon}_j)\right] \\
&= \frac{m}{m-1}\mathbb{E}\left[\varepsilon^{(1)}\varepsilon^{(1)\prime}\right] - \frac{m}{m-1}\mathbb{E}\left[\bar{\varepsilon}\bar{\varepsilon}'\right] \\
&= \frac{m}{m-1}\Sigma - \frac{1}{m-1}\Sigma = \Sigma.
\end{aligned}$$

(e) Now assume $n = d$. Is $\widehat{\Sigma}$ UMVU? Why or why not?

**Solution:**
Yes, it is. We can write the model with unknown $\Sigma$ as an exponential family with

$$\begin{aligned}
p_{\beta,\Sigma}(y) &= |2\pi\Sigma|^{-1/2}\exp\left\{\beta'X'\Sigma^{-1}y - \frac{1}{2}y'\Sigma^{-1}y - \beta'X'\Sigma^{-1}X\beta\right\} \\
&= |2\pi\Sigma|^{-1/2}\exp\left\{\left(\beta'X'\Sigma^{-1}\right)y - \frac{1}{2}\sum_{i,j}(\Sigma^{-1})_{ij}y_iy_j - A(\beta,\Sigma)\right\} \\
&= |2\pi\Sigma|^{-1/2}\exp\left\{\left(\beta'X'\Sigma^{-1}\right)y - \frac{1}{2}\sum_{i}(\Sigma^{-1})_{ii}y_i^2 - \sum_{i<j}(\Sigma^{-1})_{ij}y_iy_j - A(\beta,\Sigma)\right\},
\end{aligned}$$

so the upper triangle of $\Sigma^{-1}$, along with $\eta = \beta' X' \Sigma^{-1}$, forms a natural parameter of dimension $2n + \binom{n}{2}$, and the natural parameter space includes an open ball around $\Sigma^{-1} = I_n$ and $\eta = 0$. Hence $(\overline{Y}, \sum_k Y^{(k)}(Y^{(k)})')$ forms a complete sufficient statistic, so

$$\hat{\Sigma} = \sum_k Y^{(k)}(Y^{(k)})' - m \sum_k \overline{Y}(\overline{Y})'$$

is a function of the complete sufficient statistic. Furthermore, it is unbiased by part (d), so it is UMVU.

**2. Contamination model (25 points, 5 points / part).**

Suppose that we observe $n$ random variables $X_1, \ldots, X_n \in [0, 1]$. The observations are supposed to come from a uniform distribution, but we suspect that our sample may be contaminated by a small proportion of observations from a different, known distribution with Lebesgue density $q(x)$ ($q$ is not necessarily continuous). Assume that for some $C < \infty$, $0 \leq q(x) \leq C$ for all $x \in [0, 1]$. That is, we observe

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} p_\theta(x) = 1 - \theta + \theta q(x).$$

Assume $\theta \in [0, b]$ for some $b < 1$.

(a) Show that the maximum likelihood estimator $\hat{\theta}_n$ is consistent for the true value $\theta_0$ as $n \to \infty$.

**Solution:**
The log-likelihood for a single observation is

$$\ell_1(\theta; X_i) = \log\left(1 + \theta(q(X_i) - 1)\right) \in [\log(1 - b), \log(1 + bC - b)],$$

which is a uniformly bounded and continuous function of $\theta \in [0, b]$. As a result, by our uniform LLN, $\frac{1}{n}\ell_n(\theta; X)$ converges uniformly in probability to its expectation, which is $-D_{\text{KL}}(\theta \,\|\, \theta_0)$. Further, since the model is identifiable, the last function has a unique maximum at $\theta = \theta_0$, so by our proposition from class, the maximizer of $\ell_n(\theta; X)$ converges to $\theta_0$ in probability.

(b) Give the asymptotic distribution of the maximum likelihood estimator as $n \to \infty$, for $\theta_0 \in (0, b)$. Give an explicit expression for the asymptotic variance in terms of a definite integral (you don't need to check any regularity conditions for this part).

**Solution:**
The first two derivatives of the log-likelihood for a single observation is

$$\dot{\ell}_1(\theta; X_i) = \frac{q(X_i) - 1}{1 + \theta(q(X_i) - 1)}$$

$$\ddot{\ell}_1(\theta; X_i) = -\left(\frac{q(X_i) - 1}{1 + \theta(q(X_i) - 1)}\right)^2.$$

The asymptotic distribution of $\hat{\theta}_n$ for an interior value of $\theta$ is

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \Rightarrow N\left(0, J_1(\theta)^{-1}\right),$$

where

$$J_1(\theta) = -\mathbb{E}_\theta \ddot{\ell}_1(\theta; X_i) = \int_0^1 \left( \frac{q(X_i) - 1}{1 + \theta(q(X_i) - 1)} \right)^2 (1 + \theta(q(X_i) - 1))\, dx$$
$$= \int_0^1 \frac{(q(X_i) - 1)^2}{1 + \theta(q(X_i) - 1)}\, dx$$

(c) Find a score test for the null hypothesis that there is no contamination, against the alternative that there is some, i.e. test $H_0 : \ \theta = 0$ vs. $H_1 : \ \theta > 0$. Give an explicit expression for your test statistic and your cutoff, in terms of a definite integral and a quantile of a known distribution.

**Solution:**
The one-sided score test rejects for large values of

$$\frac{1}{\sqrt{n}} \dot{\ell}_n(0; X) = \frac{1}{\sqrt{n}} \sum_i (q(X_i) - 1) \overset{H_0}{\Rightarrow} N(0, J_1(0)),$$

where $J_1(0) = \int_0^1 (q(x) - 1)^2\, dx$, by the CLT. That is, we reject if

$$\sum_i q(X_i) > n + z_\alpha \sqrt{n J_1(0)}.$$

(d) (*) If we expand the parameter space to $[0, 1)$, is the MLE still consistent?

**Solution:**
Yes, because the log-likelihood is convex. Let $\hat{\theta}^r$ denote the restricted MLE for the model with $\theta \in \Theta^r = [0, b]$ for $b = (1 + \theta_0)/2]$, for which $\theta_0$ is in the interior of the parameter space. Then with probability approaching 1, $\hat{\theta}^r$ is in the interior of $\Theta^r$, so $\hat{\theta}^r < b$ and the log-likelihood must be decreasing on $(b, 1)$. In that case, $\hat{\theta} = \hat{\theta}^r$. Because the two estimators coincide with probability approaching 1, we have

$$\hat{\theta} - \theta_0 = (\hat{\theta} - \hat{\theta}^r) + (\hat{\theta}^r - \theta_0),$$

and both terms are approaching 0.

(e) (*) If $\theta_0 = 0$, give the distribution of the MLE as $n \to \infty$.

**Solution:**

Note that our restriction to $\theta \geq 0$ was not statistically essential: the log-likelihood would still be uniformly bounded and convex, and the parameter space compact, if we took the parameter space $[a, b]$ for any $-1/C < a < b < 1$. Hence consider an expansion of the parameter space to $[-1/2C, b]$ and let $\hat{\theta}^u$ denote the MLE for that larger model. We have

$$\sqrt{n}\hat{\theta}^u \Rightarrow N(0, J_1(0)^{-1}), \quad \text{for } J_1(0) = \int_0^1 (q(x) - 1)^2 \, dx.$$

Of course, $\hat{\theta}^u$ and $\hat{\theta}$ do not have the same distribution because $\hat{\theta}$ cannot be negative. However, because the log-likelihood is convex, we must have $\hat{\theta} = 0$ and $\hat{\theta}^u < 0$ if and only if $\dot{\ell}_n(0; X) < 0$; otherwise, $\hat{\theta} = \hat{\theta}^u \geq 0$ because $\hat{\theta}^u$ maximizes the log-likelihood over a larger parameter space. Thus, we have almost surely

$$\hat{\theta} = \max\{\hat{\theta}^u, 0\} \iff \sqrt{n}\hat{\theta} = \max\{\sqrt{n}\hat{\theta}^u, 0\}.$$

By the continuous mapping theorem, we have $\sqrt{nJ_1(0)}\hat{\theta} \Rightarrow \max\{0, Z\}$, where $Z \sim N(0, 1)$.

**3. Two-by-two count table (25 points, 5 points / part).**

Some useful facts for this problem:

- For $\theta > 0$, the Poisson density for $X \sim \text{Pois}(\theta)$ is $\frac{\theta^x e^{-\theta}}{x!}$ on $x = 0, 1, \ldots$. The mean and variance are both $\theta$.
- For $\pi_1, \ldots, \pi_d \geq 0$ with $\sum_{i=1}^d \pi_i = 1$, the multinomial density for $X \sim \text{Multinom}(n, \pi)$ is

$$p_{n,\pi}(x) = n! \cdot \prod_{i=1}^d \frac{\pi_i^{x_i}}{x_i!},$$

  on $x \in \{0, \ldots, n\}^d$ with $\sum_i x_i = n$.
- Suppose $X_i \sim \text{Pois}(\theta_i)$ with $\theta_i > 0$, independently for $i = 1, \ldots, d$, and let $X_+ = \sum_{i=1}^d X_i$ and $\theta_+ = \sum_{i=1}^d \theta_i$. Then, conditional on $X_+ = n$,

$$(X_1, \ldots, X_d) \sim \text{Multinom}(n, (\theta_1, \ldots, \theta_d)/\theta_+)$$

Assume that $X_{ij} \sim \text{Pois}(\lambda_{ij})$, independently for $i, j \in \{0, 1\}$. We will consider the model with $\lambda_{ij} = \lambda_0 \rho^{i+j}$, for $\lambda_0, \rho > 0$. Except when otherwise specified, assume both parameters are unknown.

(a) Give a complete sufficient statistic for the model and show it is complete.

**Solution:**
Let $\eta_0 = \log \lambda_0$ and $\eta_1 = \log \rho$. Then the density is

$$\begin{aligned} p(x) &= \prod_{i,j=0}^1 \frac{(\lambda_0 \rho^{i+j})^{x_{ij}} e^{-\lambda_0 \rho^{i+j}}}{x_{ij}!} \\ &= \prod_{i,j=0}^1 \exp\left\{(\eta_0 + (i+j)\eta_1)x_{ij} - e^{\eta_0 + (i+j)\eta_1}\right\} \frac{1}{x_{ij}!} \\ &= \exp\left\{\eta_0(x_{00} + x_{01} + x_{10} + x_{11}) + \eta_1(x_{01} + x_{10} + x_{11}) - A(\eta_0, \eta_1)\right\} \prod_{i,j=0}^1 \frac{1}{x_{ij}!}, \end{aligned}$$

and the natural parameters can range anywhere in $\mathbb{R}^2$. As a result, the family is full-rank and the sufficient statistic

$$T(X) = (X_{00} + X_{10} + X_{01} + X_{11}, X_{10} + X_{01} + 2X_{11})$$

is complete.

(b) Assume (for this part **only**) that $\lambda_0$ is known, but $\rho$ is unknown. Suggest a UMP test of $H_0 : \rho = \rho_0$ vs. $H_1 : \rho > \rho_0$. You do not need to give an explicit cutoff for your test but give an explicit formula for the test statistic, explain how you would find the cutoff, and explain why your test is UMP.

**Solution:**
If $\lambda_0$ is known then $\eta_0$ is known and $\eta_1$ is the only natural parameter, with corresponding sufficient statistic $T_2(X) = X_{10} + X_{01} + 2X_{11}$. As a result, the UMP test rejects when $T_2(X)$ is large. To calculate the cutoff for the test, we could simulate the data set many times under the simple null hypothesis $\theta_{ij} = \lambda_0 \rho_0^{i+j}$ for all $i, j$. If we need to we can also randomize at the boundary to make the type I error exactly $\alpha$.

(c) Assuming again that both parameters are unknown, suggest a UMPU test of $H_0 : \rho = 1$ against $H_1 : \rho > 1$. You do not need to give an explicit cutoff for your test but explain how you would calculate it. If the data are $X_{00} = X_{01} = 0$ and $X_{10} = X_{11} = 1$, calculate the (conservative, non-randomized) $p$-value for your test.

**Solution:**
In this case we are testing the null hypothesis $H_0 : \eta_1 = 0$ against the alternative hypothesis $H_1 : \eta_1 > 0$. On the null, $T_1(X) = X_{00} + X_{10} + X_{01} + X_{11}$ is complete sufficient, so we should condition on it and reject for large values of $T_2(X)$.

For this particular data set, $T_1(X) = 2$ so we observe a $\text{Multinom}(2, 1_4/4)$ distribution under the null, conditional on $T_1(X) = 2$. Our test statistic is $T_2(X) = 3$. The largest value we could have observed is $4$, if $X_{11} = 2$, which would happen with probability $(1/4)^2 = 1/16$. The second largest value we could have observed is 3, which can happen two ways: if $X_{01} = X_{11} = 1$, or if $X_{10} = X_{11} = 1$; both of these events occur with probability $2 \cdot (1/4)^2 = 1/8$. As a result, the $p$-value is

$$\mathbb{P}_{\eta_1=0}\left(T_2(X) \geq 3 \mid T_1(X) = 2\right) = \frac{1}{8} + \frac{1}{8} + \frac{1}{16} = \frac{5}{16}.$$

(d) For the same data set, $X_{00} = X_{01} = 0$ and $X_{10} = X_{11} = 1$, find the maximum likelihood estimators for $\lambda_0$ and $\rho$. Give your answers as explicit numbers.

**Solution:**

The MLE solves

$$\begin{aligned}\mathbb{E}_{\lambda_0,\rho} T_1(X) &= \lambda_0(1 + 2\rho + \rho^2) = 2 \\ \mathbb{E}_{\lambda_0,\rho} T_2(X) &= \lambda_0(2\rho + 2\rho^2) = 3\end{aligned}$$

Dividing the second equation by the first gives

$$2\hat{\rho}/(1 + \hat{\rho}) = 3/2 \Rightarrow \hat{\rho} = 3,$$

and plugging into the first equation gives $\hat{\lambda}_0 = 1/8$.

(e) (*) Now suppose we consider a relaxed model $\lambda_{ij} = f(i + j)$, for any strictly positive real-valued function $f$ on $\{0, 1, 2\}$. This includes our previous parametric model as a special case since we could have $f(i + j) = \lambda_0 \rho^{i+j}$. Does there exist a UMPU test of the null hypothesis that our previous model was correctly specified, against the alternative that it was misspecified but the relaxed model is correct? Explain why or why not. (If you say yes you only need to give enough details to establish that such a test exists).

**Solution:**
If $\zeta_k = \log f(k)$ for $k = 0, 1, 2$, then this model is also an exponential family with density

$$\begin{aligned}p(x) &= \prod_{i,j=0}^{1} \exp\left\{\zeta_{ij} x_{ij} - e^{\zeta_{ij}}\right\} \frac{1}{x_{ij}!} \\ &= \exp\left\{\zeta_0 x_{00} + \zeta_1(x_{01} + x_{10}) + \zeta_2 x_{11} - A(\zeta)\right\} \prod_{i,j=0}^{1} \frac{1}{x_{ij}!}.\end{aligned}$$

In terms of this parameterization, the null hypothesis is that $\zeta_2 - \zeta_1 = \zeta_1 - \zeta_0$, i.e.

$$H_0: \ \zeta_2 - 2\zeta_1 + \zeta_0 = 0, \quad \text{vs.} \ \ H_1: \ \zeta_2 - 2\zeta_1 + \zeta_0 \neq 0$$

Since this hypothesis sets a specific linear combination of the natural parameters to 0, and because the natural parameter space is all of $\mathbb{R}^3$, after reparameterizing the model we will simply have a two-sided test that a natural parameter of an exponential family is zero, so there will be a UMPU test given by our usual recipe.

**4. Change point problem (25 points, 5 points / part).**

Some useful facts for this problem:

- The Beta distribution Beta$(\alpha, \beta)$ with parameters $\alpha, \beta > 0$ has density

$$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad \text{where } B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

with respect to the Lebesgue measure on $(0, 1)$. Its mean and variance are

$$\mathbb{E}_{\alpha,\beta}[X] = \frac{\alpha}{\alpha+\beta}, \quad \text{Var}_{\alpha,\beta}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

- The negative binomial distribution NB$(m, \theta)$ with parameters $m \in \{1, 2, \ldots\}, \theta \in (0, 1)$ has probability mass function

$$p_{m,\theta}(x) = \binom{x+m-1}{x}\theta^x(1-\theta)^m, \quad \text{for } x \in \{0, 1, 2, \ldots\}.$$

Its mean and variance are

$$\mathbb{E}_{m,\theta}[X] = \frac{m\theta}{1-\theta}, \quad \text{Var}_{m\theta}(X) = \frac{m\theta}{(1-\theta)^2}.$$

Assume we observe independent random variables $X_i \sim \text{NB}(m, \theta_i)$ for $i = 1, \ldots, n$. Assume also that the $\theta_i$ values are constant except at some integer $k \in \{1, \ldots, n-1\}$ where they change. That is,

$$\theta_i = \begin{cases} \gamma_0 & \text{if } i \leq k \\ \gamma_1 & \text{if } i > k \end{cases},$$

for $\gamma_0, \gamma_1 \in (0, 1)$.

Until otherwise specified, assume $k$ is known. Throughout the problem, we will assume $m$ is known.

(a) Calculate the maximum likelihood estimator for $\gamma_0$ and find its asymptotic distribution if $k, n \to \infty$. You do not need to check regularity conditions.

**Solution:**
Up to terms that do not depend on $\gamma_0$ or $\gamma_1$, the log-likelihood is

$$\ell(\gamma_0, \gamma_1; X) = \sum_{i \leq k}(X_i \log\gamma_0 + m\log(1-\gamma_0)) + \sum_{i > k}(X_i \log\gamma_1 + m\log(1-\gamma_1)).$$

$$\frac{\partial}{\partial\gamma_0}\ell(\gamma_0, \gamma_1) = \frac{1}{\gamma_0}\sum_i X_i - \frac{km}{1-\gamma_0}$$

Solving for $\hat{\gamma}_0$ to zero the derivative, we get

$$\hat{\gamma}_0 = \frac{\sum_{i \le k} X_i}{km + \sum_{i \le k} X_i} = \frac{\overline{X}^{(0)}}{m + \overline{X}^{(0)}}, \quad \text{where } \overline{X}^{(0)} = \frac{1}{k} \sum_{i \le k} X_i.$$

Note that the MLE depends only on $X_1, \ldots, X_k$, so we can ignore the other $X_i$ values; the problem is identical to what we would get if we only observed $X_1, \ldots, X_k \overset{\text{i.i.d.}}{\sim} \text{NB}(m, \gamma_0)$. In that model, the Fisher information is

$$J_1(\gamma_0) = \text{Var}_{\gamma_0}(X_1/\gamma_0) = \frac{m}{\gamma_0(1-\gamma_0)^2}.$$

As a result, we have

$$\sqrt{k}\,(\hat{\gamma}_0 - \gamma_0) \Rightarrow N\left(0, J_1(\gamma_0)^{-1}\right) = N\left(0, \frac{\gamma_0(1-\gamma_0)^2}{m}\right).$$

(b) Next assume we introduce a prior distribution that $\gamma_0, \gamma_1 \overset{\text{i.i.d.}}{\sim} \text{Beta}(\alpha, \beta)$. Give the posterior distribution for $(\gamma_0, \gamma_1)$ given $X_1, \ldots, X_n$, and give the Bayes estimator for squared error loss.

**Solution:**
Ignoring factors that do not depend on $\gamma_0$ and $\gamma_1$, we have

$$\begin{aligned} p(\gamma_0, \gamma_1 \mid X) &\propto \gamma_0^{\alpha-1}(1-\gamma_0)^{\beta-1} \cdot \prod_{i \le k} \gamma_0^{X_i}(1-\gamma_0)^m \cdot \gamma_1^{\alpha-1}(1-\gamma_1)^{\beta-1} \cdot \prod_{i > k} \gamma_1^{X_i}(1-\gamma_1)^m \\ &= \gamma_0^{\alpha + \sum_{i \le k} X_i - 1}(1-\gamma_0)^{\beta+km-1} \cdot \gamma_1^{\alpha + \sum_{i > k} X_i - 1}(1-\gamma_1)^{\beta+(n-k)m-1} \\ &\propto \text{Beta}\left(\alpha + \sum_{i \le k} X_i, \beta + km\right) \times \text{Beta}\left(\alpha + \sum_{i \le k} X_i, \beta + (n-k)m\right) \end{aligned}$$

where the last expression is a product distribution over $(\gamma_0, \gamma_1)$ since they are independent.

The Bayes estimator then is

$$\hat{\gamma}_0^{\text{Bayes}} = \frac{\alpha + \sum_{i \le k} X_i}{\alpha + \sum_{i \le k} X_i + \beta + km}, \; \hat{\gamma}_1^{\text{Bayes}} = \frac{\alpha + \sum_{i > k} X_i}{\alpha + \sum_{i > k} X_i + \beta + (n-k)m}.$$

(c) (*) Find the asymptotic distribution of the Bayes estimator for $\gamma_0$, holding $\gamma_0$ and $\gamma_1$ fixed and sending $k, n \to \infty$.

**Solution:**
Again the distribution only depends on $X_1, \ldots, X_k$, so we need to find the asymptotic distribution of

$$\hat{\gamma}_0^{\text{Bayes}} = \frac{\alpha/k + \overline{X}^{(0)}}{(\alpha+\beta)/k + m + \overline{X}^{(0)}}.$$

Let $Z_k = (\alpha+\beta)/k + \overline{X}^{(0)}$; then

$$\hat{\gamma}_0^{\text{Bayes}} = \frac{Z_k}{m + Z_k} - \frac{\beta/k}{m + Z_k}. \tag{1}$$

Let $\mu = \mathbb{E}X_1 = m\gamma_0/(1-\gamma_0)$; then by Slutsky's theorem,

$$\sqrt{k}(Z_k - \mu) = \sqrt{k}(\overline{X}^{(0)} - \mu) + (\alpha+\beta)/\sqrt{k}.$$

The first term tends to $N(0, \sigma^2)$ where $\sigma^2 = \text{Var}(X_1) = m\gamma_0/(1-\gamma_0)^2$, and the second tends deterministically (and therefore in probability) to 0. Hence by Slutsky's theorem we have $\sqrt{k}(Z_k - \mu) \Rightarrow N(0, \sigma^2)$. By the delta method, then, for the differentiable function $f(z) = z/(m+z)$, we have

$$\sqrt{k}\left(f(Z_k) - f(\mu)\right) \Rightarrow N(0, \dot{f}(\mu)^2\sigma^2),$$

where $\dot{f}(z) = m/(m+z)^2$, so

$$\dot{f}(\mu)^2\sigma^2 = \left(\frac{m}{(m + m\gamma_0/(1-\gamma_0))^2}\right)^2 \frac{m\gamma_0}{(1-\gamma_0)^2} = \frac{\gamma_0(1-\gamma_0)^2}{m}.$$

Finally, the second term in (1) tends to 0 as well, since the numerator tends to zero and the denominator tends to $m + \mu > 0$, so by Slutsky's theorem we have

$$\sqrt{k}\left(\hat{\gamma}_0^{\text{Bayes}} - \gamma_0\right) \Rightarrow N\left(0, \frac{\gamma_0(1-\gamma_0)^2}{m}\right).$$

(d) Next, we relax the assumption that $k$ is known. Instead assume $n = 10$ and all we know is that $k \in \{4, 5, 6\}$. Find a minimal sufficient statistic for the three-parameter model with $\gamma_0, \gamma_1 \in (0, 1)$ and $k \in \{4, 5, 6\}$. You do not need to prove it is minimal, as long as you give the right answer.

**Solution:**
The minimal sufficient statistic is

$$T(X) = \left( \sum_{i=1}^{4} X_i, X_5, X_6, \sum_{i=7}^{10} X_7 \right).$$

To see why, note that by standard arguments $T_k(X) = \left( \sum_{i \le k} X_i, \sum_{i > k} X_i \right)$ is complete sufficient (and therefore minimal sufficient) for the submodels with $k$ fixed. That is, we must observe at least $(T_4, T_5, T_6)$ to be able to evaluate all likelihood ratios between distributions in each submodel, and $T(X)$ is equivalent to $(T_4, T_5, T_6)$, so if $T(X)$ is sufficient then it is also minimal.

Furthermore, we can also show by standard arguments that $T(X)$ is complete sufficient for the exponential family where $\theta_1 = \cdots = \theta_4$ and $\theta_7 = \cdots = \theta_{10}$, but $\theta_5$ and $\theta_6$ are unrestricted. Since our model is a submodel of that model, $T(X)$ is also sufficient for our model.

(e) Continuing with the three-parameter model above, consider a Bayesian approach where we assign priors $k \sim$ Unif$(\{4, 5, 6\})$ independently of $\gamma_0, \gamma_1 \overset{\text{i.i.d.}}{\sim}$ Beta$(\alpha, \beta)$. Give a Gibbs sampler algorithm to sample from the posterior distribution of $(k, \gamma_0, \gamma_1)$.

**Solution:**
First, we consider how to implement a Gibbs step where $k$ is held fixed and $\gamma_0$ or $\gamma_1$ is updated; this follows from part (b) where we showed

$$\gamma_0 \mid k, \gamma_1, X \sim \text{Beta}\left( \alpha + \sum_{i \le k} X_i, \beta + km \right)$$
$$\gamma_1 \mid k, \gamma_0, X \sim \text{Beta}\left( \alpha + \sum_{i > k} X_i, \beta + (n-k)m \right).$$

To update $k$, we need to first calculate

$$\begin{aligned} \omega_k(\gamma_0, \gamma_1, X) &= \prod_{i \le k} \gamma_0^{X_i} (1-\gamma_0)^m \cdot \prod_{i > k} \gamma_1^{X_i} (1-\gamma_1)^m \\ &= p(X \mid k, \gamma_0, \gamma_1) \cdot \text{const}(X) \\ &= p(X, \gamma_0, \gamma_1 \mid k) \cdot \text{const}(X, \gamma_0, \gamma_1). \end{aligned}$$

Then, we have

$$p(k \mid X, \gamma_0, \gamma_1) = \frac{p(k, X, \gamma_0, \gamma_1)}{\sum_{\tilde{k}=4}^{6} p(\tilde{k}, X, \gamma_0, \gamma_1)} = \frac{\frac{1}{3} p(X, \gamma_0, \gamma_1 \mid k)}{\sum_{\tilde{k}} \frac{1}{3} p(X, \gamma_0, \gamma_1 \mid \tilde{k})} = \frac{\omega_k}{\sum_{\tilde{k}} \omega_{\tilde{k}}}$$

Thus, to update $k$ we calculate $(\omega_4, \omega_5, \omega_6)$ and sample $k$ from a distribution proportional to that distribution. To sum up the algorithm, we can initialize $k^{(0)}$ and then take

For $t = 1, \ldots, T$ :

$$\text{Draw } \gamma_0^{(t)} \sim \text{Beta}\left(\alpha + \sum_{i \leq k^{(t-1)}} X_i, \beta + k^{(t-1)} m\right)$$

$$\text{Draw } \gamma_1^{(t)} \sim \text{Beta}\left(\alpha + \sum_{i > k^{(t-1)}} X_i, \beta + (n - k^{(t-1)}) m\right)$$

$$\text{Draw } k^{(t)} \sim q(k) \propto \omega_k \left(\gamma_0^{(t)}, \gamma_1^{(t)}, X\right).$$